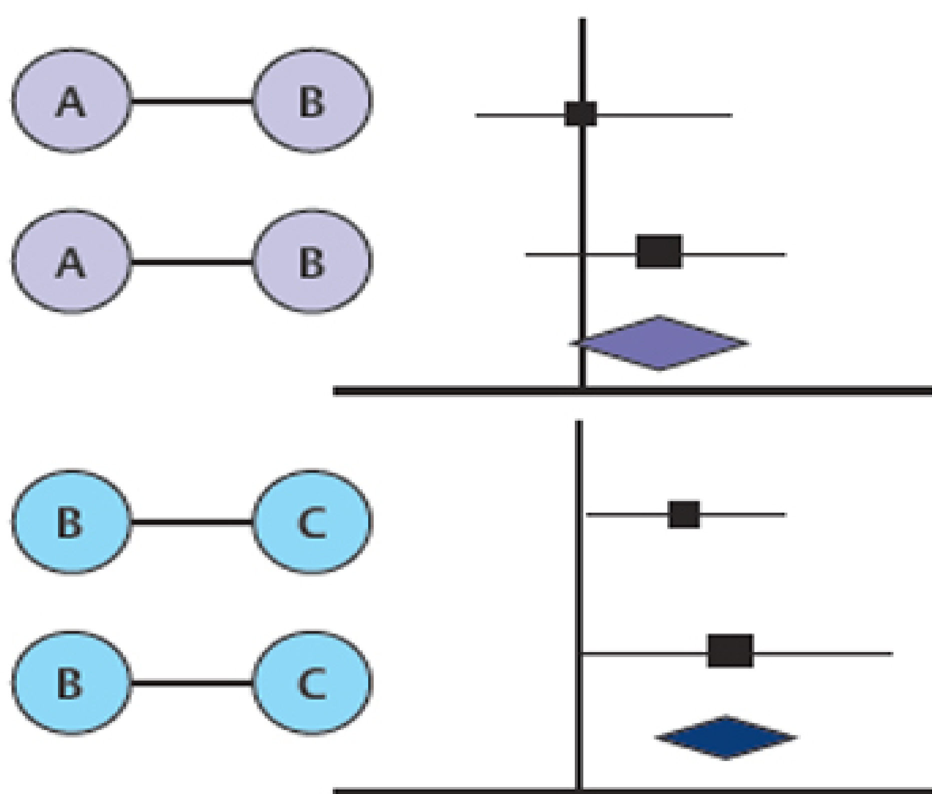
Guia Prático de Revisão Sistemática e Metanálise

Leonardo Roever





Thieme Revinter

Guia Prático de Revisão Sistemática e Metanálise

Leonardo Roever

Doutor e Mestre em Ciências da Saúde Pela Universidade Federal de Uberlândia (UFU)
Pós-Doutorando pela Universidade de São Paulo (USP)
Fundador e CEO da Rede Brasileira de Pesquisa em Meta-Análise (BRAMETIS)
Editor Associado Internacional do *European Heart Journal*
Membro do Conselho Editorial do *International Journal of Cardiology*
Editor da Seção (Medicina Baseada em Evidências, Ensaios Clínicos e suas
Interpretações – *Current Atherosclerosis Reports*
Editor Acadêmico de Medicine – Baltimore, EUA
Autor/Coautor de 201 Artigos em Periódicos Especializados – *The Lancet, The Lancet Neurology e British Medical Journal*

Thieme
Rio de Janeiro • Stuttgart • New York • Delhi

Dados Internacionais de
Catalogação na Publicação (CIP)

R719g

Roever, Leonardo
Guia Prático de Revisão Sistemática e Metanálise/Leonardo Roever – 1. Ed. – Rio de Janeiro – RJ: Thieme Revinter Publicações, 2020.

86 p.: il; 16 x 23 cm.
Inclui Índice Remissivo e Referência Bibliográfica
ISBN 978-85-5465-213-5
eISBN 978-85-5465-220-3

1. Metanálise. 2. Pesquisa. I. Título.

CDD: 610
CDU: 61

Contato com o autor:
leonardoroever@hotmail.com

Rua do Matoso, 170, Tijuca
20270-135, Rio de Janeiro – RJ, Brasil
http://www.ThiemeRevinter.com.br

Thieme Medical Publishers
http://www.thieme.com
Capa: Thieme Revinter Publicações Ltda.

Impresso no Brasil por BMF Gráfica e Editora Ltda..
5 4 3 2 1
ISBN 978-85-5465-213-5

Também disponível como eBook:
eISBN 978-85-5465-220-3

Nota: O conhecimento médico está em constante evolução. À medida que a pesquisa e a experiência clínica ampliam o nosso saber, pode ser necessário alterar os métodos de tratamento e medicação. Os autores e editores deste material consultaram fontes tidas como confiáveis, a fim de fornecer informações completas e de acordo com os padrões aceitos no momento da publicação. No entanto, em vista da possibilidade de erro humano por parte dos autores, dos editores ou da casa editorial que traz à luz este trabalho, ou ainda de alterações no conhecimento médico, nem os autores, nem os editores, nem a casa editorial, nem qualquer outra parte que se tenha envolvido na elaboração deste material garantem que as informações aqui contidas sejam totalmente precisas ou completas; tampouco se responsabilizam por quaisquer erros ou omissões ou pelos resultados obtidos em consequência do uso de tais informações. É aconselhável que os leitores confirmem em outras fontes as informações aqui contidas. Sugere-se, por exemplo, que verifiquem a bula de cada medicamento que pretendam administrar, a fim de certificar- se de que as informações contidas nesta publicação são precisas e de que não houve mudanças na dose recomendada ou nas contraindicações. Esta recomendação é especialmente importante no caso de medicamentos novos ou pouco utilizados. Alguns dos nomes de produtos, patentes e design a que nos referimos neste livro são, na verdade, marcas registradas ou nomes protegidos pela legislação referente à propriedade intelectual, ainda que nem sempre o texto faça menção específica a esse fato. Portanto, a ocorrência de um nome sem a designação de

PREFÁCIO

Desde a aurora dos tempos, as pessoas que enfrentam decisões difíceis têm confiado em suas próprias experiências, ou nas experiências de outros, dentro de um quadro mais ou menos lógico. Brutalmente falando, um experimento científico (por exemplo, um ensaio clínico) é uma coleção de experiências seguindo critérios precisos para selecionar, seguir e avaliar os assuntos. Mas cada uma dessas experiências pode ser limitada por viés sistemático ou erro aleatório. Como podemos enfrentar uma pletora conflitante de diferentes estudos enfocando o mesmo assunto?

É claro que os tomadores de decisão podem evitar decidir completamente ou usar outros fatores, como custo, apelo cultural ou preferências pessoais. Mas, se uma abordagem mais pragmática é preferida, então nos deparamos com duas alternativas principais, que são elegantemente capituladas pela Constituição dos Estados Unidos. De fato, este documento estipula que cada um dos 50 estados que constituem a União pode eleger apenas dois senadores. Assim, a tomada de decisões no Senado é baseada na contagem simples de votos, apesar de a Califórnia ser habitada por quase 40 milhões de pessoas, em desacordo com as 577.737 do Wyoming. Por outro lado, a Câmara dos Representantes possui 435 representantes, que são proporcionalmente eleitos pelos estados de acordo com o tamanho da população (por exemplo, a Califórnia elege 53, enquanto o Wyoming, apenas 1). As estipulações constitucionais visam assegurar a representatividade e reduzir o risco de que poucos estados potentes oprimam minorias ou estados minoritários, mas é evidente que a Câmara dos Representantes representa melhor toda a população dos Estados Unidos da

América. Este preâmbulo é muito importante, pois a maioria das decisões baseadas em evidências ainda é tomada usando-se uma abordagem semelhante à do Senado (por exemplo, a maioria das recomendações de diretrizes práticas), apesar de sua fraqueza evidente.

Revisões sistemáticas abrangendo metanálises (ou seja, a combinação estatística da totalidade da base de evidências sobre um determinado tópico, ponderação adequada para precisão do estudo individual) são úteis para a transição para uma abordagem semelhante à da Câmara dos Deputados, na qual pequenos estudos não são excluídos, mas recebem seu impacto adequadamente pequeno na tomada de decisão geral.

De fato, embora as eleições políticas possam parecer uma tarefa bastante simples, revisões sistemáticas e metanálises podem parecer enganosamente simples, pelo menos para o usuário comum, mas podem ser desafiadoras para projetar e conduzir, assim como interpretar. Somos, portanto, gratos a Leonardo Roever, que se comprometeu a nos fornecer um "Guia Prático de Revisão Sistemática e Metanálise" conciso e pungente, que pode oferecer uma orientação valiosa para todos os interessados e tomadores de decisão no que diz respeito à síntese de evidências. Em particular, este livro será útil para compreender os principais pontos fortes e fracos das revisões sistemáticas e metanálises, liderando o caminho para métodos mais sofisticados de síntese de evidências, como metanálise de rede, metanálise de dados de participantes individuais e revisões gerais, não obstante uma cobertura adequada de métodos e aplicações auxiliares.

Por fim, a beleza da metanálise reside em sua simplicidade, abrangência e parcimônia, que realmente brilham neste livro.

Giuseppe Biondi-Zoccai, MD, MStat
Associate Professor of Cardiology
Department of Medical-Surgical Sciences and Biotechnologies
Sapienza University of Rome
Corso della Repubblica 79
04100 Latina
Italy

giuseppe.biondizoccai@uniroma1.it

SUMÁRIO

INTRODUÇÃO

O *Guia Prático de Revisão Sistemática e Metanálise* tem como objetivo auxiliar a compreensão do processo de tomada de decisão e, com isso, aproximar o resultado das pesquisas à prática clínica. Escrito de modo claro e didático, este livro provê as informações necessárias para orientar profissionais, pesquisadores e acadêmicos a responder, com base na melhor evidência científica, a questões recorrentes na prática clínica, como a interpretação e a compreensão dos estudos de revisão sistemática e metanálise. Proporciona, ainda, a descrição da metodologia dos estudos e a avaliação de sua qualidade, facilitando ao leitor a aplicabilidade das ferramentas e dos conceitos apresentados. Original em seu formato e conteúdo, é indispensável para todo profissional atuante em saúde, seja na área de assistência, docência, pesquisa ou gestão.

Guia Prático de Revisão Sistemática e Metanálise

Thieme Revinter

REVISÃO SISTEMÁTICA

CAPÍTULO 1

RESUMO

A revisão sistemática consiste em um processo de pesquisar, selecionar, avaliar, sintetizar e relatar as evidências clínicas sobre uma determinada pergunta e/ou tópico. Nos dias atuais, a revisão sistemática é considerada uma maneira mais racional e menos tendenciosa de organizar, avaliar e integrar as evidências científicas. O objetivo deste capítulo é apresentar de forma metodológica como realizar e interpretar uma revisão sistemática.

Palavras-chave: revisão sistemática, medicina baseada em evidências, prática clínica baseada em evidências.

INTRODUÇÃO

A revisão sistemática (RS) é uma forma rigorosa de resumir as evidências científicas disponíveis que são derivadas de vários ensaios clínicos, estudos de diagnóstico e prognóstico, ou de um método em particular. Para isto, a RS utiliza uma metodologia com questões claramente desenhadas e métodos para identificar e avaliar criticamente as pesquisas de grande relevância, seguida pela organização e análise de dados dos estudos que serão incluídos na revisão. Os resultados de um único estudo podem ser aplicados a apenas certo tipo de paciente ou a um determinado ambiente clínico. Já uma RS de vários estudos sobre o mesmo tema pode fornecer informações que sejam relevantes para uma vasta

gama de pacientes em diferentes ambientes clínicos. Uma RS limita o viés dos estudos existentes, e também melhora a confiabilidade e a precisão das recomendações, por meio da combinação de informações de estudos individuais. Também possui uma dimensão da amostra total que é maior do que a de qualquer um dos estudos sobre o tema específico. No Quadro 1-1 está a *checklist* a ser incluída no relato de uma RS.[1-8]

Quadro 1-1. *Checklist* para Relato de Revisão Sistêmica

Seção/tópico	Nº	Item do ***checklist***
TÍTULO		
Título	1	Identifique o artigo como uma revisão sistemática, metanálise, ou ambos
RESUMO		
Resumo estruturado	2	Apresente um resumo estruturado incluindo, se aplicável: referencial teórico; objetivos; fonte de dados; critérios de elegibilidade; participantes e intervenções; avaliação do estudo e síntese dos métodos; resultados; limitações; conclusões e implicações dos achados principais; número de registro da revisão sistemática
INTRODUÇÃO		
Racional	3	Descreva a justificativa da revisão no contexto do que já é conhecido
Objetivos	4	Apresente uma afirmação explícita sobre as questões abordadas com referência a participantes, intervenções, comparações, resultados e desenho de estudo (PICOS)

MÉTODOS		
Protocolo e registo	5	Indique se existe um protocolo de revisão, se e onde pode ser acessado (p. ex., endereço eletrônico), e, se disponível, forneça informações sobre o registro da revisão, incluindo o número de registro
Critérios de elegibilidade	6	Especifique características do estudo (p. ex., PICOS, extensão do seguimento) e características dos relatos (p. ex., anos considerados, idioma, se é publicado) usadas como critérios de elegibilidade, apresentando justificativa
Fontes de informação	7	Descreva todas as fontes de informação na busca (p. ex., base de dados com datas de cobertura, contato com autores para identificação de estudos adicionais) e data da última busca
Busca	8	Apresente a estratégia completa de busca eletrônica para pelo menos uma base de dados, incluindo os limites utilizados, de forma que possa ser repetida
Seleção dos estudos	9	Apresente o processo de seleção dos estudos (isto é, busca, elegibilidade, os incluídos na revisão sistemática e, se aplicável, os incluídos na metanálise)
Processo de coleta de dados	10	Descreva o método de extração de dados dos artigos (p. ex., formas para piloto, independente, em duplicata) e todos os processos para obtenção e confirmação de dados dos pesquisadores

Lista dos dados	11	Liste e defina todas as variáveis obtidas dos dados (p. ex., PICOS, fontes de financiamento) e quaisquer referências ou simplificações realizadas
Risco de viés em cada estudo	12	Descreva os métodos usados para avaliar o risco de viés em cada estudo (incluindo a especificação se foi feito durante o estudo ou no nível de resultados), e como esta informação foi usada na análise de dados
Medidas de sumarização	13	Defina as principais medidas de sumarização dos resultados (p. ex., risco relativo, diferença média)
Síntese dos resultados	14	Descreva os métodos de análise dos dados e combinação de resultados dos estudos, se realizados, incluindo medidas de consistência (p. ex., I2) para cada metanálise
MÉTODOS		
Risco de viés entre estudos	15	Especifique qualquer avaliação do risco de viés que possa influenciar a evidência cumulativa (p. ex., viés de publicação, relato seletivo nos estudos)
Análises adicionais	16	Descreva os métodos de análise adicional (p. ex., análise de sensibilidade ou análise de subgrupos, metarregressão), se realizados, indicando quais foram pré-especificados
RESULTADOS		
Seleção de estudos	17	Apresente números dos estudos rastreados, avaliados para elegibilidade e incluídos na revisão, razões para exclusão em cada estágio,

		preferencialmente por meio de gráfico de fluxo
Características dos estudos	18	Para cada estudo, apresente características para extração dos dados (p. ex., tamanho do estudo, PICOS, período de acompanhamento) e apresente as citações
Risco de viés em cada estudo	19	Apresente dados sobre o risco de viés em cada estudo e, se disponível, alguma avaliação em resultados (ver item 12)
Resultados de estudos individuais	20	Para todos os resultados considerados (benefícios ou riscos), apresente para cada estudo: (a) sumário simples de dados para cada grupo de intervenção; e (b) efeitos estimados e intervalos de confiança, preferencialmente por meio de gráficos de floresta
Síntese dos resultados	21	Apresente resultados para cada metanálise feita, incluindo intervalos de confiança e medidas de consistência
Risco de viés entre estudos	22	Apresente resultados da avaliação de risco de viés entre os estudos (ver item 15)
Análises adicionais	23	Apresente resultados de análises adicionais, se realizadas (p. ex., análise de sensibilidade ou subgrupos, metarregressão [ver item 16])
DISCUSSÃO		
Sumário da evidência	24	Sumarize os resultados principais, incluindo a força de evidência para cada resultado; considere sua relevância para grupos-chave (p. ex., profissionais da

		saúde, usuários e formuladores de políticas)
Limitações	25	Discuta as limitações no nível dos estudos e dos desfechos (p. ex., risco de viés) e no nível da revisão (p. ex., obtenção incompleta de pesquisas identificadas, viés de relato)
Conclusões	26	Apresente a interpretação geral dos resultados no contexto de outras evidências e implicações para futuras pesquisas
FINANCIAMENTO		
Financiamento	27	Descreva as fontes de financiamento para a revisão sistemática e outros suportes (p. ex., suprimento de dados), papel dos financiadores na revisão sistemática

ESTRUTURA E CONTEÚDO DE UMA REVISÃO SISTEMÁTICA

Uma RS segue a mesma estrutura de um artigo original de pesquisa.

Título

O título da RS deve ser preciso e refletir claramente o tema em análise.

Resumo

Um resumo deve ser estruturado e consiste em: introdução, métodos, resultados e conclusão.

Introdução

A introdução deve resumir o tema e explicar a necessidade de realização da RS. Deve abordar as lacunas na literatura científica sobre o assunto, ou se houve divergência na literatura analisada. A introdução deve ser breve e a última sentença deve indicar claramente os objetivos da avaliação.[1-8]

Métodos

A metodologia da RS deve ser explicada de forma clara e lógica.

Critérios de Inclusão e Exclusão

Deve-se descrever os tipos de estudo que serão incluídos, por exemplo, um ensaio clínico randomizado numa determinada população para testar um determinado tipo de tratamento. Nesta seção, deve se utilizar os subtítulos, por exemplo, que tipos de paciente/pessoa/animal foram utilizados; quais intervenções/características foram procuradas; e quais as medidas dos resultados e características foram utilizadas. Deve-se mencionar algumas características-chave do estudo, tais como o tamanho da amostra, os principais resultados, e se houve restrição de idioma ou ano dentro do qual os estudos tinham que serem feitos. Quais foram os estudos excluídos da revisão?[1-8]

Identificação de Estudos

É necessário fornecer os dados sobre a pesquisa nos bancos de dados eletrônicos utilizados, incluindo as bases de dados utilizadas (p. ex., MEDLINE, Scielo, Scopus, ISI Web of Knowledge, Google Scholar e outras, se houver) e o período de pesquisa em que foram verificados os estudos. Deve-se fornecer uma lista dos termos de busca em um apêndice, e as palavras-chave utilizadas na busca de estudos devem ser listadas nesta seção.[1-8]

Na maioria dos comentários, apenas os estudos publicados no idioma inglês estão incluídos. No entanto, alguns pesquisadores podem incluir estudos em que não estão escritos na língua inglesa.

Seleção dos Estudos

Deve-se reportar os métodos de seleção e triagem de elegibilidade para todos os estudos que foram identificados, e remover os duplicados. Após isto, deve-se realizar o rastreio de relevância do título, seguido do resumo e do artigo em sua totalidade.[1-8]

Um gráfico de fluxo é recomendado para demonstrar a estratégia de busca utilizada (Fig. 1-1).

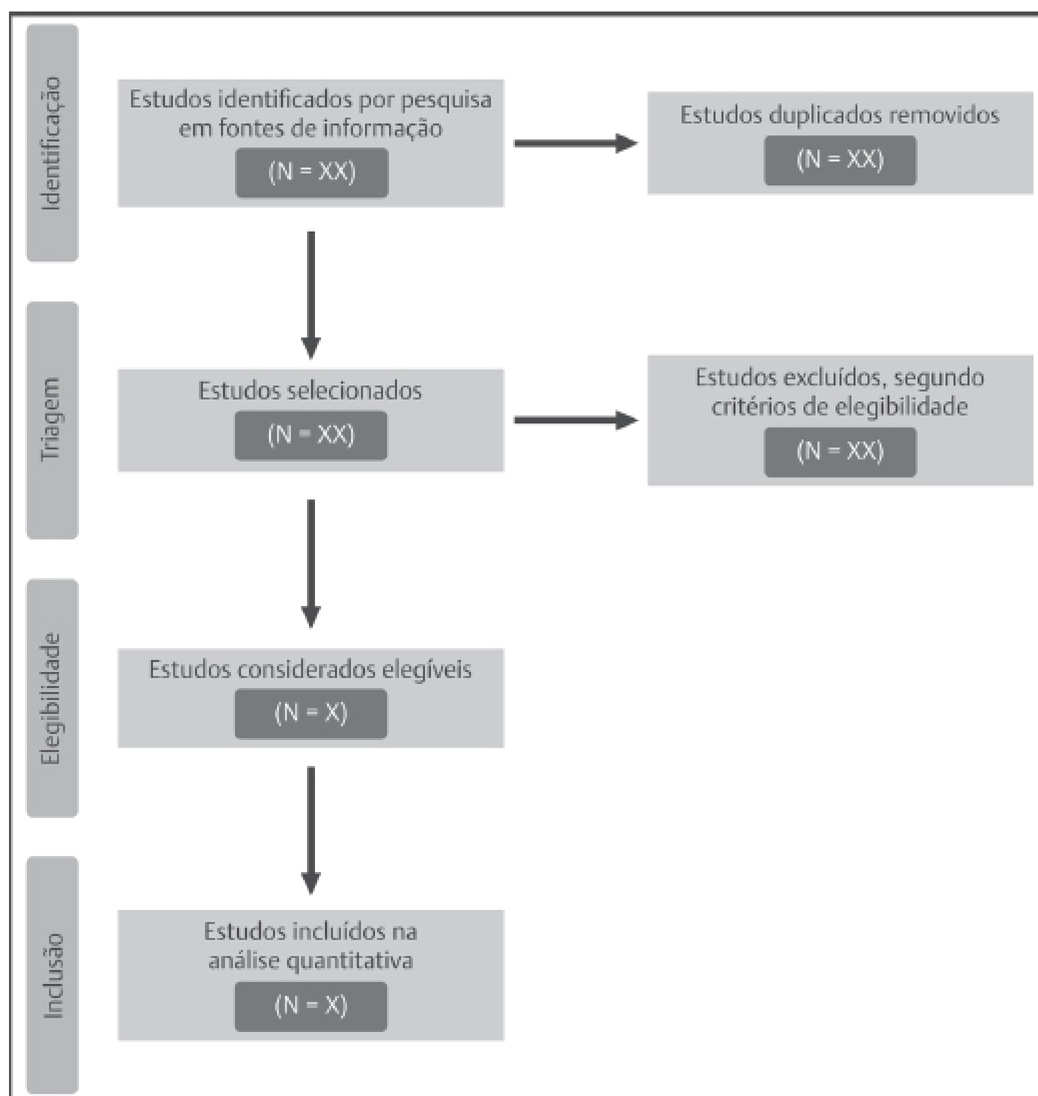


Fig. 1-1. Gráfico de fluxo para demonstrar a estratégia de busca utilizada e os estudos excluídos e incluídos.

Extração de Dados

Nesta seção, devem ser descritos quais dados foram extraídos a partir dos estudos analisados, tais como a descrição de pacientes incluídos, quantos tinham determinado resultado avaliado, o que as estatísticas sumárias demonstravam nos artigos, e se os dados foram obtidos por meio dos gráficos se estes não foram fornecidos numericamente.[1-8]

Os dados devem ser extraídos independentemente dos estudos por pelo menos dois pesquisadores para evitar tendenciosidade nas opiniões.

Avaliação da Qualidade

Os critérios de avaliação da qualidade têm sido descritos para a maioria dos tipos de estudos e muitos deles estão disponíveis no EQUATOR Network (www.equator-network.org).

Análise dos Dados

Aqui deve-se descrever sucintamente os métodos estatísticos utilizados para analisar dados. Em um comentário sobre tratamento, é típico calcular a razão de probabilidades para cada resultado com intervalos de confiança de 95% e o valor para a magnitude do efeito, e também deve-se realizar testes de heterogeneidade para verificar se os estudos eram todos provenientes e amplamente relacionados à mesma conclusão ou se houver variações significativas entre eles. A análise pode também ser influenciada pela quantidade de dados disponíveis, e com isto a heterogeneidade precisa ser avaliada para cada resultado. Os resultados podem ser classificados em diferentes subtipos e por meio disto podemos determinar se tal procedimento reduz qualquer heterogeneidade. Por exemplo, numa RS de um determinado tipo de tratamento, os estudos poderão ser classificados por doses, por tempo de administração, após o início da doença ou por tempo de avaliação dos desfechos.[1-8]

Resultados

Devem ser descritos de forma lógica e clara.

Resultados da Pesquisa

Nesta seção, deve-se mencionar o número de estudos identificados a partir de cada banco de dados analisado, quantos estudos foram duplicados, quantos foram excluídos no âmbito do inquérito do resumo, quantos artigos com o texto completo foram avaliados e, destes, quantos foram excluídos até que o número final de artigos para serem incluídos foi determinado. Também é preciso mencionar o número de artigos não incluídos na revisão devido à publicação na literatura não inglesa.[1-8]

Tipos de Estudo e Características

Aqui é preciso descrever as estatísticas demográficas e descritivas dos pacientes nos estudos selecionados. É importante indicar os tipos de paciente que foram incluídos, tais como a faixa de idade e a gravidade da doença, ou qualquer outra característica metodológica que possa influenciar as medidas dos resultados. É útil ter o cuidado de garantir que os mesmos pacientes não serão incluídos em várias publicações, o que poderia aumentar artificialmente o tamanho da amostra. Com isto, devemos reportar o número total de estudos individuais reais dentro do número total de artigos identificados. Muitas das características dos estudos podem ser resumidas nos gráficos ou tabelas.[1-8]

Qualidade dos Estudos e Fontes Potenciais

O relatório do índice médio de qualidade é derivado do método de avaliação da qualidade, e indica quais os estudos têm pontos de escore de qualidade particularmente ruim. O escore de qualidade do estudo também pode ser usado para ver se qualquer heterogeneidade aparente entre os resultados dos estudos possa ser explicada pela inclusão ou exclusão de estudos de baixa qualidade.[1-8]

Efeito da Intervenção Sobre o Resultado

Aqui devemos descrever os resultados reais da análise para estimar o total efeito entre os diferentes estudos. Como exemplo, devemos

descrever as razões de chance para cada resultado em uma revisão sobre o tratamento ou as diferenças médias em um estudo de observação. É preciso explorar as razões potenciais para a heterogeneidade e descrever o que o estudo mostrou.[1-8]

Discussão

Esta seção deve começar com uma declaração resumindo as principais conclusões da avaliação da RS. Por exemplo: são as evidências a favor do tratamento por ele ter um efeito benéfico?[1-8]

Será que os estudos sugerem que alguma característica particular está associada com risco futuro de uma doença específica? É um teste de diagnóstico melhor do que outro? Pode um procedimento intervencionista menos invasivo ser utilizado de forma confiável como uma alternativa para o procedimento invasivo convencional? Se não houver provas suficientes para chegar a uma conclusão definitiva, em seguida a declaração deve esclarecer os achados da RS.[1-8]

Os parágrafos subsequentes devem descrever as limitações dos estudos incluídos e a confiabilidade dos resultados. Quão robustos são os resultados? Houve algum viés? Após isto, podemos descrever os pontos fortes e fracos dos métodos de avaliação. Por exemplo: foram as publicações na língua não inglesa omitidas? A revisão inclui apenas um aglomerado de artigos com um pequeno tamanho da amostra? É útil colocar os resultados no contexto de outros conhecimentos sobre o tema, como exemplo por meio da comparação deste comentário com os estudos previamente publicados de revisões sistemáticas ou opiniões e diretrizes atuais. Será que a revisão destacou lacunas no conhecimento que futuros estudos devem abordar? Existe nova informação suficiente para modificar a prática clínica existente?[1-8]

E, por fim, as conclusões e as implicações para a prática atual, e particularmente para as pesquisas futuras que possam ter significativo impacto nas decisões clínicas.

REFERÊNCIAS BIBLIOGRÁFICAS

1. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ. 2009;339:b2700.
2. Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions version 5.0.2 [updated September 2011]. The Cochrane Collaboration, 2011. Disponível em www.cochrane-handbook.org.
3. Roever L, Zoccai GB. Critical Appraisal of Systematic Reviews and Meta-analyses. Evidence Based Medicine and Practice. 2015;1(1):e106.
4. Green S. Systematic reviews and meta-analysis. Singapore Med J. 2005;46:270-3.
5. Guyatt G, Drummond R. Part 1. The basics: Using the medical literature. 1A. Introduction: The philosophy of evidence based medicine. In: Guyatt G, Rennie D, eds. Users Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice. Chicago. AMA Press. 2002:3-12.
6. Ng KH, Peh WCG. Writing a systematic review. Singapore Med J. 2010;51(5):362-366.
7. Ministério da Saúde, Secretaria de Ciência, Tecnologia e Insumos Estratégicos, Departamento de Ciência e Tecnologia. Diretrizes metodológicas: elaboração de revisão sistemática e metanálise de ensaios clínicos randomizados. Brasília: Editora do Ministério da Saúde, 2012.
8. Abuabara K, Freeman EE, Dellavalle R. The Role of Systematic Reviews and Meta-analysis in Dermatology. J Invest Dermatol. 2012;132:e2.

METANÁLISE

CAPÍTULO 2

RESUMO

A metanálise é uma técnica estatística utilizada para combinar resultados provenientes de diferentes estudos. Para isto, são descritas as medidas de efeito utilizadas em metanálise na área da saúde, bem como os modelos de efeito fixo e de efeitos aleatórios O objetivo deste capítulo é apresentar de forma metodológica como realizar e interpretar uma metanálise na pesquisa cardiovascular.

Palavras-chave: metanálise; saúde baseada em evidências; metodologia.

INTRODUÇÃO

A metanálise é utilizada para combinar resultados provenientes de diferentes estudos e, com isto, produzir estimativas que resumem o todo, denominadas estimativas metanalíticas. Para se considerar que o resultado de uma metanálise tenha significado aplicado, os estudos que compõem os seus dados devem ser o resultado de uma revisão sistemática, que consiste de um conjunto de regras para identificar os estudos sobre uma determinada questão, e então selecionar quais deles serão incluídos. No Quadro 2-1, estão os elementos de uma revisão sistemática.[1-14]

Quadro 2-1. Elementos de uma Revisão Sistemática

1	Definir a questão clínica

2	Identificar todos os estudos completos sobre a questão, publicados ou não
3	Selecionar os estudos que preencham padrões elevados de validade científica
4	Procurar por evidências de viés nos estudos selecionados
5	Descrever a qualidade científica dos estudos
6	Questionar se a qualidade está sistematicamente relacionada aos resultados do estudo
7	Descrever os estudos com uma figura (gráfico de floresta)
8	Decidir se os estudos são suficientemente semelhantes para justificar uma combinação
9	Se eles forem suficientemente semelhantes para serem combinados, calcular uma medida sumária de efeito e seu intervalo de confiança

O objetivo deste capítulo é apresentar os conceitos básicos e a forma metodológica de como realizar uma metanálise sem a utilização de fórmulas complexas, o que facilitará a compreensão do profissional de saúde.

COMPOSIÇÃO DOS DADOS

Primeiramente, é necessário definir quais os resultados serão combinados para que os estudos possam ser combinados por meio de uma metanálise. Estes estudos podem comparar diferentes tecnologias (drogas, dispositivos, tratamentos, procedimentos etc.) e apresentam medidas do tamanho do efeito com a razão de chances, o risco relativo, a redução absoluta de risco e o número necessário para tratar.[1,12-14]

MEDIDAS DE EFEITO

Para que os estudos possam ser combinados em uma metanálise, é preciso escolher uma medida de efeito e calculá-la para cada estudo. Estas medidas são classificadas em quatro categorias: (I) medidas de efeito para descrever apenas um grupo com uma só variável binária, (II) medidas de efeito para comparar dois grupos por meio de variáveis binárias, (III) medidas de efeito para comparar dois grupos por meio de variáveis contínuas, e
(IV) medidas de efeito para grupos com dados categóricos ordinais.[1,12-14]

Medidas de Efeito para um Grupo

São utilizadas quando o interesse é descrever um estudo ou estimar um efeito. As medidas de efeito mais utilizadas para um grupo são o *odds* e a incidência. As estimativas destas medidas são calculadas e seu uso em metanálise requer também uma medida de variabilidade.[1,12-14]

Odds *ou Chances*

Por definição, *odds* é a razão entre a probabilidade de o evento ocorrer e a probabilidade do evento não ocorrer, o que é equivalente à razão entre o número de eventos e o número de não eventos.[1,12-14]

Incidência

A incidência mede o número de novos casos surgidos numa determinada população e num determinado intervalo de tempo.[1,12-14]

Medidas de Efeito para Comparação de Dois Grupos por Meio de Variáveis Binárias

Os dados provenientes de estudos que comparam dois grupos quanto a uma variável binária (ocorrência ou não de um evento de interesse) geralmente são organizados em tabelas 2 × 3, conforme o Quadro 2-2.[5]

As medidas de efeito mais comuns calculadas para este tipo de dados são *odds ratio*, risco relativo, redução absoluta de risco e número necessário para tratar.[1,12-14]

Quadro 2-2. Apresentação dos Dados dos Estudos

Dados do estudo	Evento	Não evento	Tamanho dos grupos
Grupo 1	A	B	AB
Grupo 2	C	D	CD

Odds *Ratio ou Razão de Chances*

O *odds ratio* expressa quantas vezes o *odds* de um grupo equivale ao *odds* do outro grupo. Um *odds ratio* igual a 1 indica que não existe diferença entre os *odds* dos dois grupos, um *odds ratio* menor que 1 indica que o grupo 1 tem *odds* menor que o grupo 2, e um *odds ratio* maior que 1 indica o contrário.[1,12-14]

Odds *Ratio de Peto ou Razão de Chances de Peto*

O *odds ratio* de Peto é uma forma alternativa para se estimar o *odds ratio*. A ideia e a interpretação são as mesmas, mas uma diferença relevante é que, no *odds ratio* de Peto, a, b, c ou d podem ser iguais a zero. O *odds ratio* de Peto é particularmente útil quando os tamanhos das amostras dos grupos não são muito diferentes.[1,12-14]

Risco Relativo

O risco relativo é definido como a razão entre a probabilidade do evento ocorrer no grupo 1 e a probabilidade do evento ocorrer no grupo 2. Um risco relativo igual a 1 indica que os dois grupos apresentam o mesmo risco, um risco relativo menor que 1 indica que o grupo 1 tem risco menor que o grupo 2, e um risco relativo maior que 1 indica o contrário.[1,12-14]

Redução Absoluta de Risco

A redução absoluta de risco é definida como a diferença entre as probabilidades de ocorrer o evento nos dois grupos.[1,12-14]

Número Necessário para Tratar

O número necessário para tratar (NNT) é o número de indivíduos que devem ser submetidos ao tratamento a fim de prevenir um evento. O NNT é dado pelo inverso da redução absoluta de risco.[1,12-14]

Medidas de Efeito para Comparação de Dois Grupos por Meio de Variáveis Contínuas

Quando os dados dos estudos são medidos em uma escala contínua, é comum aparecerem médias e desvios-padrão para resumir as informações. Utilizando média e desvio-padrão, as medidas de efeito mais comuns para dados contínuos são a diferença absoluta entre médias e a diferença padronizada entre médias.[1,12-14]

Diferença Absoluta entre Médias

A diferença absoluta entre as médias quantifica a diferença entre a média do grupo 1 e a média do grupo 2. Esta medida deve ser utilizada quando todos os estudos da metanálise utilizam a mesma escala contínua.[1,12-14]

Diferença Padronizada entre Médias

A diferença padronizada entre médias pode ser utilizada quando os dados são contínuos, mas usam diferentes escalas. Existem três maneiras diferentes de se estimar a diferença padronizada entre médias. São elas: d de Cohen, g ajustado de Hedges, e Δ de Glass. Estes métodos diferem quanto ao uso do desvio-padrão nos cálculos e quanto ao uso ou não de correção para evitar vícios quando as amostras são pequenas (Quadro 2-3).[1,12-14]

Quadro 2-3. Maneiras Diferentes de Se Estimar a Diferença Padronizada entre Médias

d de Cohen	O d de Cohen é a diferença média de cada estudo dividida pelo desviopadrão, criando, assim, uma medida-padrão que possa ser comparável entre os estudos
g ajustado de Hedges	O g ajustado de Hedges é similar ao d de Cohen, porém ele inclui um ajuste para corrigir o vício de amostras pequenas
Δ de Glass	O Δ de Glass utiliza o desvio-padrão do grupo controle para dividir a diferença de médias

MODELOS DE EFEITO FIXO E ALEATÓRIOS

Em metanálise são utilizados basicamente dois tipos de modelos: os modelos de efeito fixo e os modelos de efeitos aleatórios.[1,12-14]

Modelo de Efeito Fixo

Pressupõe que o efeito de interesse é o mesmo em todos os estudos e que as diferenças observadas entre eles são devidas à variabilidade interna de cada estudo.[1,12-14]

Modelo de Efeitos Aleatórios

Pressupõe que o efeito de interesse não é o mesmo em todos os estudos. Estes estudos são parte de uma amostra aleatória de uma população hipotética de estudo. Apesar dos efeitos dos estudos não serem considerados iguais, eles são conectados por meio de uma distribuição de probabilidade, geralmente suposta como normal.[1,12-14]

É utilizado quando se deseja combinar os dados que são parecidos ao ponto de ser coerente juntar a informação deles em uma medida resumo. Mas existem razões para que os estudos não sejam considerados idênticos quanto ao efeito do tratamento, quando, por exemplo, o interesse for por estudos em que está sendo testada a eficácia de uma droga, exercício ou dispositivo. Entre eles, pode haver diferença nos grupos de pessoas

selecionadas, um grupo do estudo pode ser mais saudável, a faixa etária dos pacientes pode variar de estudo para estudo, a classe social pode ser diferente, entre outros fatores que podem influenciar para que o efeito do tratamento seja diferente entre os estudos. Quando isto acontece, estas diferenças devem ser investigadas e alternativas como uma metanálise em subgrupos ou uma metarregressão devem ser consideradas. A alternativa nesta situação é a utilização de um modelo de efeitos aleatórios que considere a existência não apenas da variação dentro de cada estudo, mas também a variação entre os estudos. Ou seja, tal modelo pressupõe que os efeitos dos estudos não são iguais, mas que são conectados por meio de uma distribuição de probabilidade, geralmente suposta como normal.[1,12-14]

Método de Mantel-Haenszel

Este método foi proposto para combinar *odds ratio* oriundos de diferentes estratos quando não são detectadas diferenças significativas entre eles, e é utilizado para combinar medidas de efeito oriundas de diferentes estudos quando a suposição é de que o efeito é o mesmo em todos os eles, ou seja, a mesma suposição do modelo de efeito fixo. Quando comparado com o modelo de efeito fixo, o método de Mantel-Haenszel é mais robusto quando as amostras forem pequenas, visto que, nestas situações, produz estimativas melhores para os erros-padrão das medidas de efeito.[1,12-14]

Escolha do Modelo

O modelo de efeito fixo é adequado quando acreditamos que o efeito do tratamento é idêntico entre os estudos, quando o objetivo for estimar um efeito do tratamento em uma população específica e não extrapolar para outras populações.[1,12-14]

O modelo de efeitos aleatórios pode ser utilizado caso combine vários estudos que têm o mesmo objetivo e que não foram conduzidos da mesma maneira. Neste caso, é possível extrapolar para outras populações, o que torna a análise mais abrangente. Um ponto que deve ser verificado é que, se o número de estudos for

muito pequeno, a estimativa da variância entre eles não terá uma boa precisão. Neste caso, devem-se relatar os efeitos separados e não como uma medida resumo, utilizar um modelo de efeito fixo e não fazer inferências para outras populações. Podemos também utilizar a abordagem bayesiana, onde a estimativa da variância entre os estudos pode ser baseada também em dados externos aos estudos utilizados. Quanto à heterogeneidade, ela ajuda na escolha do modelo, mas não é determinante.[1,12-14]

ABORDAGEM BAYESIANA

O método bayesiano na metanálise é uma técnica recente que ainda vem sendo explorada. As suas vantagens básicas são: (I) a possibilidade de considerar o conhecimento prévio,
(II) a flexibilidade da modelagem, e (III) a naturalidade da interpretação dos resultados. No enfoque da metanálise, quando o número de estudos selecionados na revisão sistemática é pequeno, têm-se dificuldades de estimar as medidas de efeito metanalíticas por intermédio da teoria clássica, pois esta depende de pressupostos assintóticos. Nestes casos, a metanálise bayesiana também é sugerida como uma possível solução para obtenção de estimativas combinadas mais confiáveis. O Quadro 2-4 mostra breve comparação entre inferência clássica e inferência bayesiana.[15,16]

Sendo assim, a estatística bayesiana constitui-se em um processo de diminuição de incertezas sobre o desconhecido que se baseia em dados estatísticos e em evidências prévias.

METANÁLISE DE SUBGRUPOS

A metanálise é utilizada para se estimar o efeito médio do tratamento; mas, por mais bem-feita que seja a revisão sistemática, ainda pode haver diferenças significativas entre os estudos: tratamento aplicado, qualidade do controle, estimador utilizado etc. Estas diferenças devem não apenas ser consideradas no planejamento da metanálise, mas também explicadas.[1,12-16]

Quadro 2-4. Comparação entre a Inferência Clássica e a Inferência Bayesiana[15,16]

Inferência clássica	Inferência bayesiana
O parâmetro desconhecido é considerado uma constante fixa	O parâmetro desconhecido é considerado uma variável aleatória que segue uma distribuição a *priori*
É considerada somente a informação amostral por meio da verossimilhança	Podem-se considerar, a *priori*, as informações de estudos anteriores, conhecimento pessoal etc.
Não se pode falar em probabilidade para as estimativas dos intervalos de confiança	Pode-se falar em probabilidade para as estimativas dos intervalos de credibilidade

A metanálise de subgrupos é feita por meio da distinção de grupos de estudo. É um método de segmentação de tal modo que exista grande homogeneidade dentro de cada grupo e heterogeneidade entre eles. É feita uma comparação entre os estudos individuais que têm um maior poder estatístico de detectar uma possível existência de subgrupos com efeitos diferentes não detectados pelos estudos individuais, ou ainda refutar a alegação de efeitos dependentes de subgrupos. No entanto, o tamanho amostral em análise de subgrupos é frequentemente menor e, por isso, estas análises geralmente perdem em poder estatístico.[1,12-16]

Num conjunto de estudos clínicos no qual exista a intervenção de diferentes drogas nos grupos de tratamento, o objetivo é realizar uma metanálise para a medida de efeito, mas deve-se considerar previamente uma análise de agrupamento para que esta heterogeneidade seja explicada. Com isto, deve-se definir o modelo de efeito a ser considerado em cada subgrupo: efeito fixo (quando não há heterogeneidade significativa dentro do subgrupo) ou efeitos aleatórios (quando há heterogeneidade). Nestes modelos, o efeito

do tratamento é avaliado no subgrupo definido ou num subgrupo complementar.[1,12-16]

Na análise bayesiana, uma das vantagens sobre o enfoque clássico é a possibilidade de estimar o chamado intervalo de credibilidade, que é fornecido por meio da distribuição *a posteriori,* e outra vantagem agregada à distribuição *a posteriori* é o cálculo de probabilidades. Com isto, é possível saber qual a probabilidade do *odds* tratamento ser maior que o *odds* controle, ou seja, além da estimativa pontual e do intervalo de credibilidade, também temos como resultados as probabilidades de ocorrências do *odds ratio*. Portanto, no modelo de análise de subgrupos sob enfoque bayesiano, estas vantagens são estendidas para a diferença entre os subgrupos, e com isto pode-se obter a estimativa para o intervalo de credibilidade e a probabilidade da diferença entre os subgrupos, o que não é possível no enfoque clássico.[1,12-16]

Modelo de Efeito Fixo em Metanálise de Subgrupos

O modelo de efeito fixo em metanálise de subgrupos possui as mesmas considerações vistas anteriormente. As diferenças entre os estudos são devidas apenas à variabilidade interna de cada um, isto é, a variabilidade entre os estudos é considerada nula.[1,12-16]

Modelo de Efeitos Aleatórios em Metanálise de Subgrupos

O modelo de efeitos aleatórios considera a variação dentro de cada estudo e a variação entre os estudos. Pode-se especificar o modelo de efeitos aleatórios para subgrupos da forma de um modelo hierárquico.[1,12-16]

METARREGRESSÃO

A metarregressão é feita para se verificar se as características dos estudos possibilitam uma síntese das informações e permitem explicar a heterogeneidade possivelmente existente entre estudos.

A variável dependente é o tamanho do efeito e as covariáveis são as características presentes nos estudos. Por exemplo, idade, proporção de indivíduos, duração dos estudos, número de terapêuticas etc. Na metarregressão, o número de covariáveis a incluir está limitado ao número de estudos considerados na metanálise. Idealmente deve ser utilizada uma covariável por cada 10 estudos.[1]

HETEROGENEIDADE

Em estudos de metanálise, quando consideramos que a variabilidade entre os estudos não é apenas aleatória, dizemos que os estudos são heterogêneos. A avaliação sobre heterogeneidade tem papel fundamental na escolha do modelo de metanálise, e deve ser realizada antes da escolha do modelo.[1,12-16]

As maneiras mais utilizadas de se verificar a existência de heterogeneidade em metanálises são pelo teste Q de Cochran ou pela estatística I^2 de Higgins e Thompson. Quando nos referimos à heterogeneidade em medidas de efeito, pensamos na sua verdadeira variação; entretanto, essa variação é constituída da verdadeira heterogeneidade e também do erro aleatório.[1,12-16]

Separação da Variação

Os efeitos observados variam pela heterogeneidade na medida de efeito e também pelo erro aleatório dentro do estudo. Para quantificar a heterogeneidade, é necessário separar a variação nestes dois componentes.[1,12-16]

Para separar estas variações, é necessário calcular a variação total observada de estudo a estudo e estimar o quanto os efeitos observados devem variar supondo-se homogeneidade entre os estudos. Então, calcula-se o excesso de variação pela diferença da variação total pela variação supondo homogeneidade, pois esta é a heterogeneidade.[1,12-16]

Teste Q de Cochran

O teste Q de Cochran apresenta como hipótese nula a afirmação de que os estudos são homogêneos. Uma deficiência deste teste é possuir um baixo poder quando o número de estudos que compõem a metanálise é pequeno. Por outro lado, quando o número de estudos é muito grande, ele pode detectar uma falsa heterogeneidade.[1,12-16]

Estatística I^2

A estatística I^2 foi proposta como uma forma mais direta de verificar a existência de heterogeneidade por meio de proporção. Quando o valor for negativo, ele é igualado a 0. A significância de I^2 é equivalente à significância de Q. Na estatística I^2, uma escala em que um valor de I^2 é próximo de 0% indica não heterogeneidade entre os estudos, próximo a 25% indica baixa heterogeneidade, próximo a 50% indica heterogeneidade moderada, e próximo a 75% indica alta heterogeneidade entre os estudos.[1,12-16]

FOREST PLOT

O *forest plot* ou gráfico de floresta é um gráfico que mostra informações individuais de cada estudo incluído na metanálise e também da medida metanalítica. Para cada estudo, o gráfico apresenta a medida de efeito e seu intervalo de confiança, sendo que a medida de efeito é representada por um símbolo que pode ser um quadrado, um círculo, ou outra figura dependendo do *software*. O tamanho deste símbolo é proporcional ao peso do estudo na metanálise: quanto maior o peso, maior o tamanho do símbolo da medida de efeito. Também é exibida em torno da medida de efeito uma linha horizontal que é o seu intervalo de confiança, sendo que, quanto maior for esta linha, maior será a variabilidade dentro do estudo.[1,12-17]

CONCLUSÃO

Este capítulo abordou as questões básicas de uma metanálise, como as medidas de efeito, os modelos de efeito fixo e de efeitos

aleatórios, como identificar heterogeneidade, a análise bayesiana e de subgrupos entre os estudos, e de como escolher o modelo a ser utilizado na área da saúde. A compreensão desta metodologia e a interpretação da metanálise é de fundamental importância nas pesquisas na área da saúde.

REFERÊNCIAS BIBLIOGRÁFICAS

1. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. Introduction to Meta-Analysis. Chichester: John Wiley & Sons; 2009.
2. Brockwell SE, Gordon R. A comparison of statistical methods for meta-analysis. Statistic in Medicine. 2001;20(6):825-840.
3. Rodrigues CL, Ziegelmann KP. Metanálise: um guia prático. Rev HCPA. 2010;30(4):436-447.
4. Egger M, Smith GD, Altman DG. Systematic reviews in health care: Meta-analysis in context. BMJ Books; 2001.
5. Fletcher RH, Fletcher SW. Epidemiologia clínica: elementos essenciais. Porto Alegre: Artmed; 2013.
6. Glass GV. Primary, secondary, and meta-analysis of research. Educational Researcher. 1976;5(10):3-8.
7. HedgeS LV, Olkin I. Statistical Methods for Meta-Analysis. San Diego, CA: Academic Press; 1985.
8. Kulinskaya E, Morgenthaler S, Staudte RG. Meta-Analysis: A guide to calibrating and combining statistical evidence. Chichester: John Wiley & Sons; 2008.
9. Leandro G. Meta-analysis in medical research: The handbook for the understanding and practice of meta-analysis. London: BMJ Books; 2005.
10. Martinez EZ. Metanálise de ensaios controlados aleatorizados: aspectos quantitativos. Rev Med. (Ribeirão Preto) 2007;40(2):223-235.
11. Normand SLT. Meta-analysis: formulating, evaluating, combining, and reporting. Statistics in Medicine. 1999;18:321-359.
12. Sutton AJ, Abrams KR, Jones DR, Sheldon TA, Song F. Methods for Meta-Analysis in Medical Research. Chichester:

John Wiley & Sons; 2000.
13. Whitehead A. Meta-analysis of controlled clinical trials. Chichester: John Wiley & Sons; 2002.
14. Berman NG, Parker RA. Meta-analysis: neither quick nor easy. BMC Medical Research Methodology. 2002;2:10.
15. Ziegelmann PK. Inferência Bayesiana e Teoria das Decisões. Porto Alegre: Universidade Federal do Rio Grande do Sul. Instituto de Matemática. Departamento de Estatística; 2010.
16. Ziegelmann PK. Metanálise MTC: O uso combinado de evidência direta e indireta na comparação de múltiplos tratamentos. In: 19º SINAPE, 2010 Julho 26-30. São Pedro, SP, Brasil.
17. Ziegelmann PK. Aspectos Conceituais e Práticos sobre Metanálise. Rev Enferm UFPI. 2013;2(Esp.):8-13.

METANÁLISE DE REDE

CAPÍTULO 3

RESUMO

A metanálise de rede (MAR) é uma técnica utilizada para comparar múltiplos tratamentos simultaneamente em uma única análise combinando evidências diretas e indiretas dentro de uma rede de ensaios randomizados controlados. A MAR pode ajudar a avaliar a eficácia comparativa de diferentes tratamentos regularmente utilizados na prática clínica. O objetivo deste trabalho é ilustrar o processo de metanálise de rede. Para isto, discutimos os pontos-chave da MAR, bem como as considerações para o desenvolvendo de questões de pesquisa apropriadas, a condução da revisão da literatura, o resumo de dados, as sínteses qualitativa e quantitativa, a apresentação dos resultados, o desenho da discussão e as conclusões de uma MAR.

Palavras-chave: metanálise de rede; comparações indiretas; comparações de tratamento misto; comparação de tratamento múltiplo; transitividade.

INTRODUÇÃO

As revisões sistemáticas de ensaios controlados randomizados (ECRs) são geralmente consideradas o maior nível de evidência para a eficácia relativa das intervenções.[1,2] Metanálise é uma técnica estatística para síntese quantitativa de estudos similares de

uma revisão sistemática.[1-3] A abordagem de metanálise convencional é útil, mas é limitada na medida em que só pode comparar duas intervenções ao mesmo tempo.[4-8] Neste capítulo, fornecemos um tutorial e discutiremos aspectos importantes da MAR.

A MAR amplia os princípios de metanálise para a avaliação de múltiplos tratamentos em uma única análise. Isto é conseguido combinando as evidências direta e indireta. A evidência direta refere-se às evidências obtidas por ECRs. Por exemplo, em um julgamento comparando os tratamentos A e B, a evidência direta é a estimativa dos efeitos relativos entre A e B. A evidência indireta refere-se à evidência obtida por meio de um ou mais comparadores comuns. Por exemplo, na ausência de ECRs que avaliam diretamente as intervenções A e B, A e B podem ser comparadas indiretamente se ambas tiverem sido comparadas com C nos estudos (formando um “laço” de A-B-C de evidências). A combinação de evidências diretas e indiretas é chamada de evidência mista.[1,4-8] A aplicação de MAR para comparações indiretas e mistas é utilizada na determinação de quanto o tratamento B é melhor do que A e de quanto o tratamento C é melhor do que o mesmo comparador A, não existindo estudos que comparem diretamente B e C (Fig. 3-1).

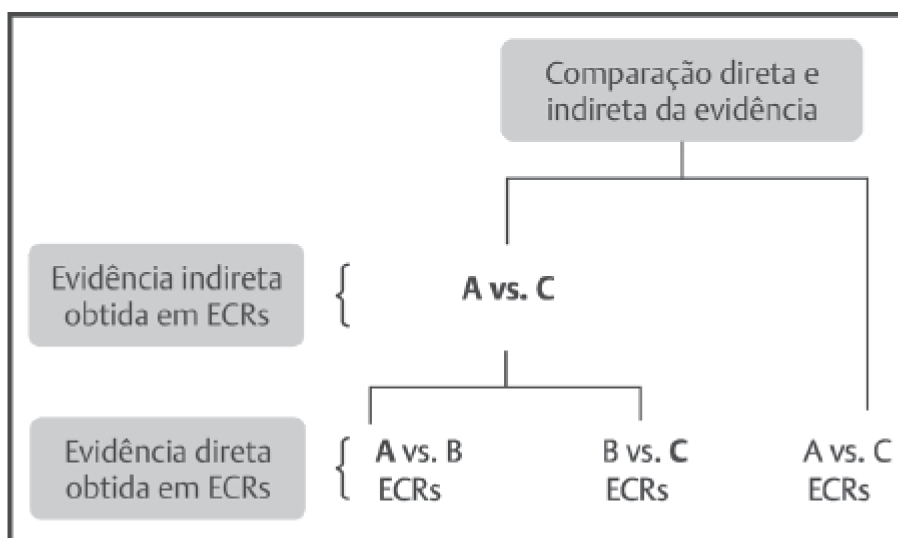


Fig. 3-1. Ilustração de uma metanálise de rede que combina a evidência direta obtida nos ECRs (A *vs.* B, B *vs.* C e A *vs.* C) e a evidência indireta obtida por meio de ECRs com o uso de um comparador comum (A *vs.* B e B *vs.* C).

Para permitir comparações dos efeitos do tratamento em todas as intervenções na metanálise de rede, a base de evidências utilizada para a metanálise deve corresponder a uma rede conectada. Em termos simples, isso significa que qualquer um de dois tratamentos pode ser comparado diretamente (cabeça a cabeça) ou indiretamente por meio de um (ou mais) referente intermediário comum.[2]

A Figura 3-2 representa quatro redes de evidências conectadas. Os pontos representam intervenções, e as linhas (isto é, conexões) implicam que um ou mais ECR comparou os respectivos tratamentos diretamente. Quaisquer dos tratamentos que não foram comparados cabeça a cabeça são passíveis de uma comparação indireta. Por exemplo, a rede na Figura 3-2a inclui estudos AB, que comparam a intervenção B com estudos A, e os estudos AC, que comparam a intervenção C com A.[2] O efeito relativo do tratamento de C *versus* B pode ser obtido com a comparação indireta dos estudos AB e AC. A rede na Figura 3-2b também inclui estudos de CD, e pode informar sobre qualquer comparação emparelhada (isto é, contrastes) entre A, B, C e D. Nas redes da Figura 3-2c, d têm um circuito fechado, o que implica que, para algumas das comparações de tratamento, há evidências diretas e indiretas.[2] Na Figura 3-2c, há evidências diretas e indiretas para as comparações AB, AC e BC. Na Figura 3-2d, há evidências diretas e indiretas de todas as comparações, com exceção do contraste de AD e BC, para o qual há apenas evidências indiretas.[2]

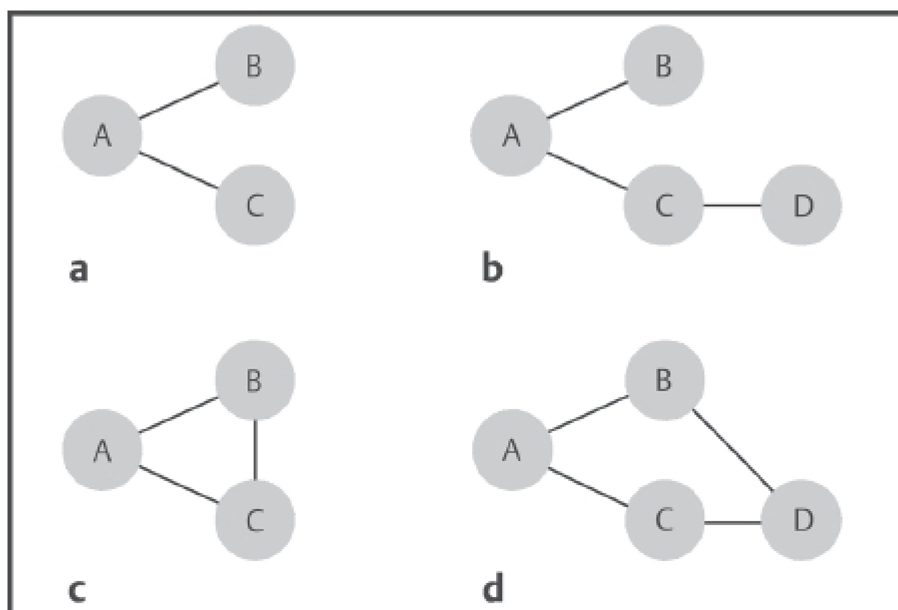


Fig. 3-2. Redes conectadas de ensaios controlados randomizados para permitir a metanálise de rede e comparações indiretas. Todas as intervenções que podem ser indiretamente comparadas são parte de uma rede.

O conceito e as definições, bem como o resumo das etapas para as principais diferenças na revisão sistemática convencional e metanálise, e as considerações adicionais para MAR são mostrados no Quadro 3-1.[5]

Quadro 3-1. Conceito, Definições e Resumo das Etapas de uma Revisão Sistemática Convencional e Considerações Adicionais para uma Metanálise de Rede[5]

Passos de uma revisão sistemática	**Considerações para uma rede de metanálise**
1. Defina a questão da revisão e o critério de elegibilidade	A questão deve se beneficiar da metanálise de rede Definir a rede de tratamento
2. Procurar e selecionar os estudos	Certificar-se de que a pesquisa seja ampla o suficiente para capturar tratamentos de interesse

3. Resumir os dados e a avaliação do risco de viés	Resumir as informações sobre possíveis modificadores de efeitos que podem violar a suposição de transitividade
4. Sintetizar evidências qualitativamente	Avaliar a geometria da rede Avaliar a transitividade
5. Sintetizar a evidência quantitativamente	Conduzir as metanálises em dois pares primeiro Usar modelos estatísticos apropriados para a metanálise de rede Avaliar a inconsistência Resumir os resultados usando abordagens adequadas para a metanálise de rede, como gráficos de rede Avaliar a superfície sob as curvas de classificação cumulativa (SUCRA)
6. Comparador comum	É a âncora à qual as comparações do tratamento estão apoiadas. Se uma rede tiver três tratamentos (A, B e C) e A estiver diretamente ligado a B enquanto C também está diretamente ligado a B, o comparador comum dessa rede é B
7. Comparação de tratamento direto (CTD)	É a estimativa usando-se comparações separadas de duas intervenções (p. ex., A contra B; B versus C) e levando-se em consideração um comparador comum (neste caso, B). Assim, os efeitos de tratamento direto de cada intervenção contra o comparador comum são usados para estimar uma evidência indireta entre as duas intervenções

8. Metanálise de rede (MAR) ou Comparação de tratamento misto (CTM)/metanálise (CTm)	Esses termos referem-se a situações que envolvem a comparação simultânea de três ou mais intervenções. Todos os tratamentos MAR que consistem em laços estritamente fechados podem ser considerados como uma série de CTDs. Na CTM, as informações diretas e indiretas estão disponíveis para informar as estimativas do tamanho do efeito para pelo menos algumas das comparações. Visualmente, isso é mostrado por laços fechados em um gráfico de rede. Os laços fechados não são obrigados a estar presentes para cada comparação em estudo. MAR é um termo inclusivo que incorpora os cenários de comparações de tratamento indireto e misto
9. Diagrama de rede e geometria	A partir da análise de rede, é gerado um diagrama de rede (gráfico) onde cada ponto representa uma intervenção e as linhas de conexão entre elas representam um ou mais ECRs nos quais as intervenções foram comparadas diretamente. A descrição das características das redes de intervenções pode incluir o uso de estatísticas de um resumo numérico, e é considerada uma avaliação da geometria da rede
10. Circuito fechado	No diagrama de rede, cada comparação tem evidências diretas e indiretas. Por exemplo, a comparação de BC tem provas

	diretas dos ensaios BC e provas indiretas dos ensaios AB e AC (e de forma semelhante para as comparações AB e AC)
11. Ordem de classificação ou rankeograma	Cálculos da probabilidade de um tratamento ser o melhor, segundo melhor, e assim por diante, para um resultado específico
12. Inconsistência ou incoerência	Conflitos estatísticos no modelo de rede (em relação à fonte de evidência, grau de similaridade de dados, falta de informação consistente) que devem ser investigados para garantir a robustez do modelo

Uma MAR válida irá satisfazer a hipótese de transitividade, que não existem diferenças sistemáticas entre as comparações disponíveis além dos tratamentos comparados. Outra maneira de ver isso é que, em um ECR hipotético composto por todos os tratamentos incluídos na MAR, os participantes podem ser randomizados para qualquer um dos tratamentos.[6-10] Por exemplo, em determinada doença, determinado tratamento é prescrito como monoterapia para o tratamento inicial, e as terapias combinadas são usadas apenas em pacientes cuja doença é insuficientemente controlada pela monoterapia.[6-10]

Uma MAR válida irá satisfazer a hipótese de transitividade, ou seja, que não existem diferenças sistemáticas entre as comparações disponíveis além dos tratamentos comparados.[9] Por exemplo, em um ECR composto por todos os tratamentos incluídos na MAR, os participantes podem ser randomizados para qualquer um dos tratamentos.[9] Em outro exemplo, em determinada doença é utilizada a monoterapia para se realizar o tratamento e as terapias combinadas são usadas apenas em pacientes cuja doença não está suficientemente controlada.[11] Portanto, os ECRs que examinam o

tratamento de primeira linha não incluem terapias combinadas e, incluindo tratamentos combinados em uma MAR de tratamentos de primeira linha, irão introduzir intransitividade.

Outro exemplo em que a transitividade seria violada é quando o tratamento para uma condição é baseado em biomarcadores genéticos. Nestes casos, os genes são usados para determinar o plano de tratamento. Como o tratamento-alvo de um gene é usado para se realizar a terapia, vamos supor que este é gene-positivo, mas não gene-negativo. Portanto, tratamentos para o gene-positivo e para o gene-negativo não seriam avaliados no mesmo ECR, e não devem ser incluídos na mesma MAR. A avaliação do pressuposto de transitividade é crítica porque a existência de intransitividade será um viés de estimativas de efeito de tratamento.[12]

DEFINIR A QUESTÃO DE PESQUISA E A REDE DE TRATAMENTO

Tal como acontece com qualquer revisão sistemática, o primeiro passo é definir a questão da pesquisa. A abordagem recomendada para o desenvolvimento da questão da pesquisa é o PICO (isto é, participantes, intervenção, comparador e resultado).[12] A MAR pode ser usada para responder pela eficácia comparativa de questões de pesquisa nas quais intervenções múltiplas estão disponíveis e podem ser usadas para uma determinada condição clínica.[13]

O próximo passo para a MAR é definir a rede de tratamento, sendo necessário tomar decisões sobre o tamanho da rede e quão distintamente os tratamentos devem ser examinados. Deve-se ter cuidado para incluir todos os tratamentos relevantes para a população ter uma base de evidências completas; mas, quanto maior as redes, maiores serão os recursos necessários para adquirir os dados, e pode ser mais provável que isto viole a transitividade.[14-16]

As intervenções como o placebo ou nenhum tratamento são geralmente consideradas mesmo que não sejam de interesse clínico porque elas podem informar outras comparações fornecendo evidência indireta.[10] A identidade distinta das intervenções avaliadas

também é importante, pois estas podem ser divididas em drogas individuais ou doses específicas de drogas, ou agrupadas em classes de drogas ou tipo de tratamento (p. ex., físico, dieta, dispositivos, comportamental etc.).[10] A decisão de dividir ou agrupar intervenções deve se dar com base na relevância clínica. Se todos os medicamentos dentro de uma classe forem considerados clinicamente intercambiáveis, então pode ser suficiente para examinar os tratamentos como classes.[9] Se houver questões importantes sobre a classe, a droga e até mesmo nível de dose, os autores podem definir redes separadas para análise.[10]

PROCESSO DE COLETA DE DADOS

A coleta de dados na metanálise consiste em uma pesquisa na literatura científica, triagem de estudos e abstração dos dados como pré-especificado em um protocolo.[17] Para a abstração de dados, é importante separar a informação sobre os modificadores dos efeitos potenciais dos estudos para permitir avaliação da transitividade, pois estes modificadores dos efeitos clínicos e as características metodológicas dos estudos incluídos podem afetar o tamanho do efeito.[17] Os modificadores de efeito devem ser pré-especificados no protocolo com base na experiência clínica ou em uma revisão da literatura anterior.[17] As características geralmente relevantes são os critérios de elegibilidade do estudo, as características da população, o *design* do estudo e o risco de viés.[17]

Análise dos Dados

Antes da análise dos dados, é importante compreender a geometria da rede, pois esta irá mostrar quais intervenções foram comparadas diretamente nos ECRs e que só podem ser informadas indiretamente.[18] A geometria da rede pode ser visualizada usando-se um gráfico de rede (Fig. 3-3).

SÍNTESE QUALITATIVA

Toda revisão sistemática deve conduzir uma síntese qualitativa antes da síntese quantitativa, e isso se aplica igualmente para a MAR. Na MAR, a síntese qualitativa inclui a avaliação das heterogeneidades clínica e metodológica, como em uma revisão sistemática convencional, bem como a transitividade.

Se houver uma variação substancial nos modificadores de efeitos potenciais entre os estudos, isso sugere a presença de heterogeneidade importante. Se existe uma variação substancial entre comparações sobre modificadores de efeito, isto sugere intransitividade. Por exemplo, todos os estudos para uma comparação podem se concentrar em populações não obesas jovens, enquanto os estudos para todas as outras comparações examinam principalmente adultos obesos mais velhos. Este cenário é problemático quando a idade e o índice de massa corporal é um modificador de efeito. Se a intransitividade é suspeitada com base na síntese qualitativa, então uma síntese quantitativa pode não ser apropriada.

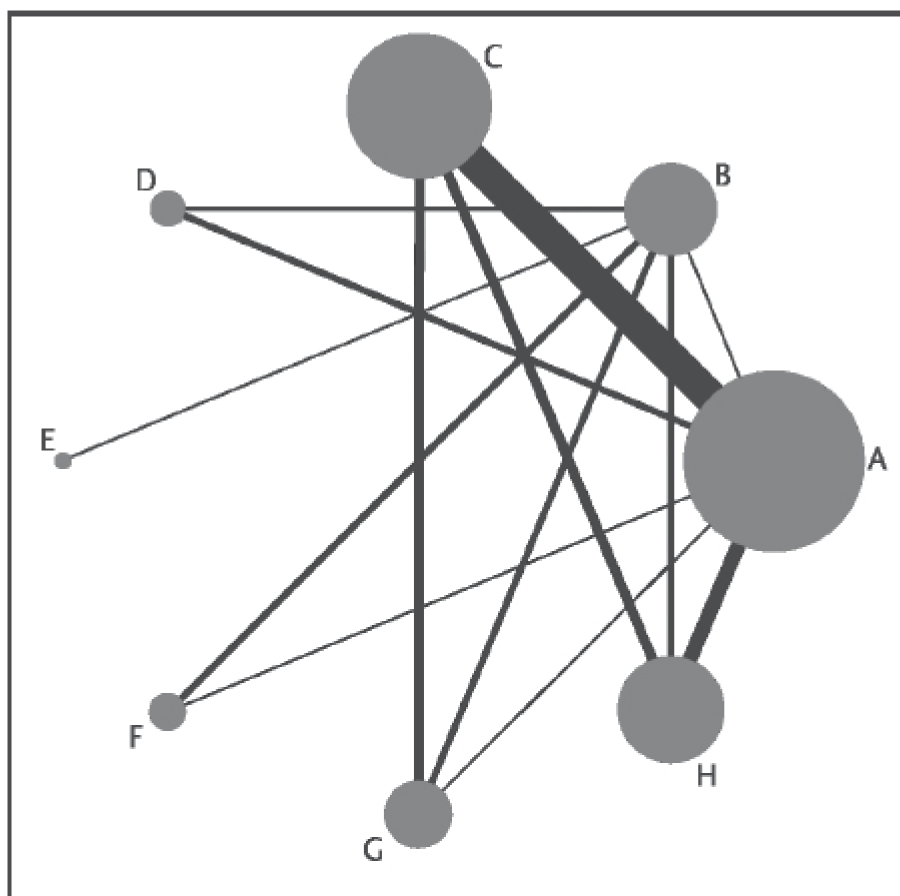


Fig. 3-3. A largura das setas e o tamanho dos pontos podem mostrar a quantidade de informações disponíveis. Por exemplo, podem ser desenhados proporcionalmente ao número de testes, número de participantes ou precisão.[18]

SÍNTESE QUANTITATIVA

A metanálise emparelhada de todas as intervenções diretamente comparadas deve ser realizada antes de se conduzir a MAR, de modo que a heterogeneidade estatística dos estudos em cada comparação possa ser avaliada.[18] A alta heterogeneidade estatística em metanálises emparelhadas pode afetar a confiança dos resultados da MAR.

O próximo passo é desenvolver o modelo MAR. Dentre os modelos comumente usados, está o modelo multivariado ou o modelo hierárquico.[19-23] Em ambos os casos, o tratamento de referência é comparado com todos os tratamentos na análise. A referência selecionada geralmente é um grupo de placebo ou

nenhum tratamento, ou o tratamento comparador mais comum. Todos os outros tratamentos podem ser comparados entre si por meio da referência.

Também é importante especificar como a heterogeneidade será assumida como ativa no modelo MAR. A heterogeneidade pode ser comparativa específica ou comum em comparações. Muitas MARs assumem uma heterogeneidade comum quando há apenas um pequeno número de estudos por comparação direta, uma vez que a estimativa de heterogeneidade pode ser mais poderosa ao tomar força nas comparações.[2]

Por outro lado, heterogeneidades específicas de comparação podem ser preferidas quando muitos estudos informam cada comparação e a suposição de uma heterogeneidade comum é não plausível.[6]

Uma outra consideração a ser tomada na MAR é a consistência estatística de acordo com comparações diretas e indiretas. A consistência é a manifestação estatística da transitividade para os dados. As abordagens para verificar a inconsistência podem ser classificadas em duas categorias: as abordagens globais e as abordagens locais. Nas abordagens globais, a inconsistência é avaliada em toda a rede, modificando o modelo MAR para explicar a incoerência potencial; ao passo que as abordagens locais detectam potenciais *hot spots* de inconsistência na rede, como, por exemplo, examinando separadamente os *loops* de evidência. Em geral, é recomendável usar ambos os tipos de método para inconsistência. A inconsistência pode ser verificada usando rotinas nos códigos Stata ou R ou WinBUGS.[24-26]

Se a inconsistência for identificada, existem algumas abordagens para lidar com isso. Primeiro, os revisores devem certificar-se de que não há erros na extração de dados. Em seguida, é importante examinar mais de perto os potenciais modificadores de efeitos de estudos dentro de laços inconsistentes. O modelo de metarregressão de rede poderia ser ajustado para verificar como os potenciais modificadores de efeito podem afetar os resultados.

As análises de sensibilidade excluindo estudos que podem ser fontes de inconsistência também podem ser úteis para avaliar a robustez dos resultados. Se houver uma inconsistência substancial

e as fontes não puderem ser identificadas, a MAR pode não ser o método apropriado para sintetizar os dados.[6]

APRESENTAÇÃO DOS RESULTADOS

Estimativas de Efeito Relativo

A MAR permite comparar o efeito relativo entre qualquer par de intervenções.

Podemos apresentar esses resultados usando uma matriz quadrada chamada liga. A tabela de liga contém todas as informações sobre eficácia relativa e sua incerteza quanto a todos os possíveis pares de intervenções (Fig. 3-4).

Por exemplo, a primeira célula no canto superior esquerdo mostra que a droga significativamente tem efeito benéfico em comparação com o placebo.

Probabilidades de Classificação

Uma das vantagens da MAR é que permite a classificação das intervenções. Com base nos resultados da MAR, podemos calcular a probabilidade de cada intervenção ter um grau específico.[4]

É desencorajado classificar os tratamentos com base na probabilidade de ser considerado como o melhor porque essa abordagem não explica a incerteza nas estimativas de efeito relativo e no ranking relativo. É mais apropriado inferir sobre o *ranking* do tratamento usando o grau médio ou as probabilidades cumulativas de classificação.[4]

Tratamento A			
OR, 1,28 (95% CrI, 0,48-3,45)	Tratamento B		
OR, 1,96 (95% CrI, 0,99-4,48)	OR, 1,53 (95% CrI, 0,62-4,32)	Tratamento C	
OR, 2,36 (95% CrI, 1,22-4,63)	OR, 1,84 (95% CrI, 0,74-4,66)	OR, 1,20 (95% CrI, 0,60-2,12)	Controle ou placebo

Fig. 3-4. representação da tabela de liga. or: razão de chances.

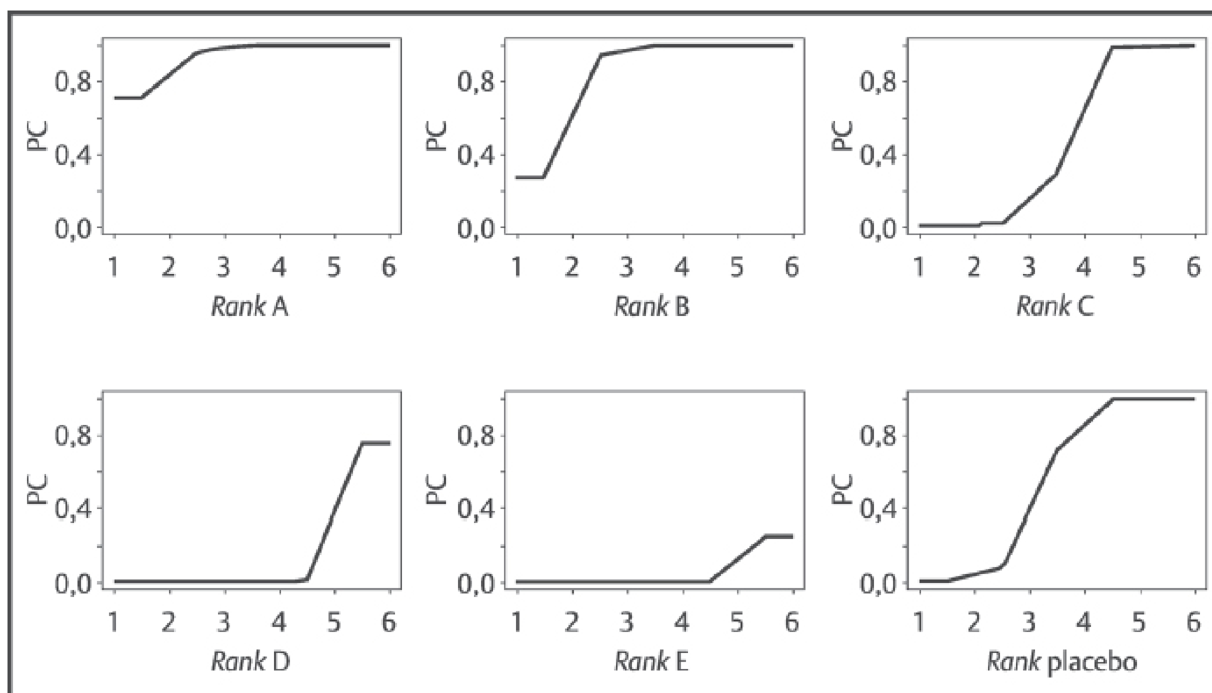


Fig. 3-5. Gráfico de SUCRA. PC: probabilidade cumulativa.

A Figura 3-5 mostra as curvas cumulativas de probabilidade de classificação para cada tratamento em nossa rede. Usando esses gráficos, podemos classificar os tratamentos de acordo com a superfície sob as curvas de classificação cumulativa (SUCRA). O valor SUCRA representa a probabilidade de um tratamento ser uma das melhores opções.[4]

Um valor de SUCRA de 100% indica que o tratamento é seguro para ser o mais efetivo na rede, enquanto um valor de 0% indica que ele será o menos efetivo. Quanto maior o valor do SUCRA, melhor será o grau de intervenção na rede.[4]

INTERPRETANDO OS RESULTADOS E FAZENDO AS CONCLUSÕES

Muito cuidado deve ser tomado ao interpretar os resultados e tirar conclusões da MAR. As conclusões extraídas da MAR devem

depender dos resultados analisados, da utilidade clínica das intervenções que pode ser mais compreendida se a análise considerar a eficácia e os resultados de segurança.[4] Também é importante considerar se os resultados avaliados são resultados substitutos ou resultados clínicos, e se são importantes para o paciente.[4] Na presença de heterogeneidade ou inconsistência na rede, falta de qualidade dos estudos subjacentes ou apenas uma pequena quantidade de dados disponíveis, os resultados precisam ser interpretados com cautela.[4]

O *ranking*, embora atraente para facilitar as decisões clínicas, pode ser enganador. Mesmo quando são utilizadas estatísticas de classificação adequadas, como os valores de SUCRA, os tratamentos de alta qualidade ainda podem ter efeitos clínicos modestos ou insignificantes e, portanto, devem ser interpretados no contexto dos seus efeitos.[27] Se o método SUCRA produzir uma classificação distinta para cada tratamento, mas as diferenças de efeito de tratamento dentro da classe forem pequenas e potencialmente não clinicamente significativas, indicando que fatores diferentes da eficácia, tais como custo, efeitos colaterais e preferências do paciente, isto pode ser importante na seleção do tratamento.[4]

Uma abordagem comum para avaliar a qualidade das evidências a partir de uma revisão sistemática é a classificação da avaliação de avaliações, desenvolvimento e avaliação (GRADE).[28] Na abordagem GRADE, a evidência é avaliada com base em seis domínios: limitações do estudo, heterogeneidade e inconsistência, indireção, imprecisão e viés de publicação.[28]

As considerações para as limitações do estudo incluem o desenho do estudo (randomizado *versus* não randomizado) e o risco de viés dos estudos individuais. Nos estudos não randomizados ou randomizados, significativas heterogeneidade e inconsistência irão diminuir a evidência.[4]

A indireção refere-se à aplicabilidade do conjunto de evidências à questão do interesse em termos de população, intervenções e resultados. Se a revisão estiver interessada em considerar pacientes com comorbidades, mas os testes apenas incluem

pacientes sem comorbidades, a evidência seria rebaixada para a indireção.[4] A evidência também seria rebaixada por indireção se os ensaios só examinassem os resultados de substituição, em vez de resultados clínicos ou importantes do paciente de interesse para a revisão. Se as estimativas de efeito tiverem amplos intervalos de incerteza cruzando ou perto do efeito nulo, a evidência é rebaixada para imprecisão.[4]

Finalmente, evidências de viés de publicação, que os estudos têm maior probabilidade de ser publicados se tiverem resultados favoráveis para o tratamento de interesse, irão diminuir a evidência. O viés de publicação pode ser suspeitado se a evidência direta provém de um pequeno número de estudos financiados comercialmente ou com base em avaliação quantitativa utilizando parcelas de funil com contorno reforçado.[4,29,30]

Existem duas abordagens propostas para aplicar GRADE à NMA.[30,31] Ambas as abordagens começam por avaliar cada domínio para cada comparação direta. Nesta abordagem de Salanti *et al*., as classificações em todos os domínios são combinadas para conferir classificações específicas de domínio, e as classificações específicas de domínio são combinadas para determinar a qualidade da evidência indireta ou mista global.[30]

Em Puhanet *et al*., a abordagem das classificações para os seis domínios são combinadas para cada comparação para fornecer classificações específicas de resumo de comparação e, em seguida, as classificações de resumo são combinadas para determinar a qualidade das evidências indiretas e mistas.[31] As conclusões da MAR podem ser usadas para informar as diretrizes de prática clínica porque as diretrizes geralmente visam fazer recomendações de tratamento no contexto de várias opções disponíveis.[32]

Relatório de Metanálise de Rede

Para tornar os resultados de uma revisão sistemática útil para pesquisadores, clínicos, tomadores de decisão e pacientes, é importante garantir que todo o processo seja transparente para que os pontos fortes e fracos da avaliação possam ser devidamente

avaliados.[4] Para este fim, os itens de Relatórios Preferenciais para Análises Sistemáticas e Metanálises (PRISMA) foram desenvolvidos para fornecer orientações sobre pontos essenciais para informar a transparência.[4] Devido aos aspectos únicos da MAR, uma extensão da declaração PRISMA para NMA foi publicada recentemente.[33,34]

Os itens adicionados à declaração de extensão incluem a apresentação, o resumo e a avaliação da geometria da rede, e a avaliação e análise da inconsistência. As modificações importantes da declaração original incluem a descrição dos critérios de elegibilidade para os tratamentos e a forma como os tratamentos serão agrupados ou divididos usando abordagens específicas da MAR para apresentar resultados (como tabelas da liga e *rankings* de tratamento) e discutindo a validade do pressuposto de transitividade.[4]

Precauções sobre Metanálise de Rede

A MAR pode ser uma ferramenta poderosa para a pesquisa de eficácia comparativa devido à sua capacidade de avaliar indiretamente evidências e classificar os tratamentos. No entanto, a MAR é mais complexa do que a metanálise em pares, devendo a suposição de transitividade ser rigorosa e considerada ao longo de todo o processo de MAR. Análises adicionais, como a metarregressão da rede, muitas vezes são necessárias e aumentam ainda mais a complexidade da análise.[4]

Além disso, a MAR é muito intensiva em recursos. Como as MARs geralmente fazem perguntas mais amplas, muitas vezes elas envolvem mais estudos em cada etapa da revisão sistemática, desde a triagem até a análise, do que a metanálise convencional. Antes de realmente realizar um MAR, é importante antecipar o tempo e o compromisso de recursos necessários.[4,35-39]

CONCLUSÃO

Em resumo, a MAR é um método promissor que pode informar a pesquisa de eficácia comparativa na presença de múltiplos tratamentos, permitindo incrementar o processo de síntese da

melhor evidência disponível e auxiliar a tomada de decisões. A questão clínica deve ser desenvolvida com a contribuição de um especialista clínico da área do assunto e um estatístico com conhecimento em MAR. As avaliações de transitividade e consistência, assim como um bom relatório, são fundamentais para garantir que a MAR seja válida.

REFERÊNCIAS BIBLIOGRÁFICAS

1. Roever L, Biondi-Zoccai G. Network Meta-analysis to Synthesize Evidence for Decision Making in Cardiovascular Research. Arq Bras Cardiol. 2016;106(4):333-7.
2. Biondi-Zoccai G. Network meta-analysis: evidence synthesis with mixed treatment comparison. Hauppauge, NY: Science Publishers; 2014.
3. Biondi-Zoccai G, Abbate A, Benedetto U, Palmerini T, D'Ascenzo F, Frati G. Network meta-analysis for evidence synthesis: What is it and why is it posed to dominate cardiovascular decision making? Int J Cardiol. 2015;182:309-314.
4. Rouse B, Chaimani, Li T. Network meta-analysis: an introduction for clinicians. Intern Emerg Med. 2017;12(1):103-111.
5. Greco T, Biondi-Zoccai G, Saleh O, Pasin L, Cabrini L, Zangrillo A, et al. The attractiveness of network meta-analysis: a comprehensive systematic and narrative review. Heart Lung Vessel. 2015;7(2):133-142.
6. Tonin FS, Rotta I, Mendes AM, Pontarolo R. Network meta-analysis: a technique to gather evidence from direct and indirect comparisons. Pharm Pract (Granada). 2017;15(1):943.
7. Cipriani A, Higgins JPT, Geddes JR, Salanti G. Conceptual and technical challenges in network meta-analysis. Ann Intern Med. 2013;159(2):130-137.
8. Roever L. Understanding meta-analysis studies in clinical research. Rev Soc Bras Clin Med. 2016;4(4):245-9.
9. Salanti G. Indirect and mixed-treatment comparison, network, or multiple-treatments meta-analysis: many names, many benefits,

many concerns for the next generation evidence synthesis tool. Res Synth Methods. 2012;3(2):80-97.
10. Li T, Puhan MA, Vedula SS et al. Network meta-analysis highly attractive but more methodological research is needed. BMC Med. 2011;9:79.
11. Lindsley K, Yu T, Li T, Michelessi M. Combination medical treatment for primary open angle glaucoma and ocular hypertension: a network meta-analysis. Cochrane Database Syst Rev. 2018;(5):CD011366.
12. Yu Q, Zhu Z, Liu Y, Zhang J, Li K. Efficacy and safety of HER2-targeted agents for breast cancer with HER2-overexpression: a network meta-analysis. PLoS One. 2015;10(5):e0127404.
13. Russell R, Chung M, Balk EM, Atkinson S, Giovannucci EL, Ip S, et al. Issues and challenges in conducting systematic reviews to support development of nutrient reference values: workshop summary. Agency for Healthcare Research and Quality (US), Rockville; 2009.
14. Caldwell DM. An overview of conducting systematic reviews with network meta-analysis. Syst Rev. 2014;3:109.
15. Mavridis D, Giannatsi M, Cipriani A, Salanti G. A primer on network meta-analysis with emphasis on mental health. Evid Based Ment Heal. 2014;18(2):40-46.
16. Sturtz S, Bender R. Unsolved issues of mixed treatment comparison meta-analysis: network size and inconsistency. Res Synt Methods 2012;3:300-311.
17. Li L, Tian J, Tian H, Moher D, Liang F, Jiang T, et al. Network meta-analyses could be improved by searching more sources and by involving a librarian. J Clin Epidemiol. 2014;67:1001-1007.
18. Mills EJ, Thorlund K, Ioannidis JP. Demystifying trial networks and network meta-analysis. BMJ Br Med J. 2013;2914:10-15.
19. Deeks JJ, Higgins JPT, Altman DG. Analysing data and undertaking meta-analyses. In: Higgins JPT, Green S, eds. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011). The Cochrane Collaboration; 2011.

20. Higgins JPT, Jackson D, Barrett JK, Lu G, Ades AE, White IR. Consistency and inconsistency in network meta-analysis: concepts and models for multi-arm studies. Res Synth Methods. 2012;3(2):98-110.
21. White IR. Network meta-analysis. Stat J. 2015;15(4):951-985.
22. Salanti G, Higgins JP, Ades A, Ioannidis JP. Evaluation of networks of randomized trials. Stat Methods Med Res. 2008;17(3):279-301.
23. Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. Stat Med. 2004;23(20):3105-3124.
24. Dias S, Welton NJ, Sutton AJ, Caldwell DM, Lu G, Ades AE. Evidence synthesis for decision making 4: inconsistency in networks of evidence based on randomized controlled trials. Med Decis Making. 2013;33(5):641-656.
25. White IR, Barrett JK, Jackson D, Higgins JPT. Consistency and inconsistency in network meta-analysis: model estimation using multivariate meta-regression. Res Synth Methods. 2012;3(2):111-125.
26. Song F, Xiong T, Parekh-Bhurke S, Loke YK, Sutton AJ, Eastwood AJ, et al. Inconsistency between direct and indirect comparisons of competing interventions: meta-epidemiological study. BMJ. 2011;343:d4909.
27. Rouse B, Cipriani A, Shi Q, Coleman AL, Dickersin K, Li T. Network meta-analysis for clinical practice guidelines—a case study on first-line medical therapies for primary open-angle glaucoma. Ann Intern Med. 2016;164(10):674-682.
28. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ. 2008;336(7650):924-926.
29. Guyatt GH, Oxman AD, Montori V, Vist G, Kunz R, Brozek J, et al. GRADE guidelines: 5. Rating the quality of evidence—publication bias. J Clin Epidemiol. 2011;64(12):1277-1282.
30. Salanti G, Del Giovane C, Chaimani A, Caldwell DM, Higgins JP. Evaluating the quality of evidence from a network meta-analysis. PLoS One. 2014;9(7):e99682.

31. Puhan MA, Schunemann HJ, Murad MH, et al. A GRADE Working Group approach for rating the quality of treatment effect estimates from network meta-analysis. BMJ. 2014;349:g5630.
32. Leucht S, Chaimani A, Cipriani AS, Li T, Brignardello-Petersen R, Singh JA, et al. Network meta-analysis should be the highest level of evidence in treatment guidelines. Eur Arch Psychiatry Clin Neurosci. 2016;(6):477-480.
33. Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement (Reprinted from Annals of Internal Medicine). PLoSMed. 2009;6(7):e1000097.
34. Hutton B, Salanti G, Caldwell DM, Chaimani A, Schmid CH, Cameron C, et al. The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: checklist and explanations. Ann Intern Med. 2015;162(11):777-84.
35. Roever L, Resende ES, Biondi-Zoccai G, Roever-Borges AS. Degrees of Recommendations and Levels of Evidence: What you Need to Know? Evidence based Medicine and Practice. 2016;2(2):101.
36. Roever L, Biondi-Zoccai G. Critical Appraisal of a Network Meta-Analysis in Emergency Medicine. Austin Emerg Med. 2016;2(7):1037.
37. Roever L, Resende ES, Diniz ALD, Penha-Silva N, Biondi-Zoccai G, Casella-Filho A, et al. Critical analysis of clinical research articles: a guide for evaluation. Evidence based Medicine and Practice. 2016;2:e116.
38. Roever L. Understanding Systematic Review Studies. Rev Soc Bras Clin Med. 2017;15(2):127-130.
39. Roever L, Biondi-Zoccai G. Understanding Umbrella Review of the Cardiovascular Research. Int J Cardiovasc Sci. 2016;29(6):500-503.

CAPÍTULO 4

METANÁLISE INDIVIDUAL DE DADOS DE PARTICIPANTES

RESUMO

Metanálise é o método estatístico utilizado na revisão sistemática para integrar os resultados dos estudos incluídos e aumentar o poder estatístico da pesquisa primária. Estudos de metanálise decorrentes de uma revisão sistemática envolvem a combinação e análise de evidências que são utilizadas para produzir resultados baseados em um conjunto de pesquisas prévias. Os métodos tradicionais de metanálise sintetizam os dados agregados obtidos de publicações de estudo como uma estimativa de efeito de tratamento (*odds ratio*, risco relativo) e sua incerteza associada (erro-padrão ou intervalo de confiança). Uma abordagem alternativa é a metanálise individual de dados de participantes ou de pacientes na qual os dados de nível individual bruto para cada estudo são obtidos e utilizados para síntese. O objetivo deste capítulo é apresentar de forma metodológica como realizar e interpretar uma revisão sistemática e uma metanálise individual de dados de participantes.

Palavras-chave: metanálise; saúde baseada em evidências; metodologia.

INTRODUÇÃO

O termo “dados de participantes ou pacientes individuais” refere-se aos **dados relacionados para cada participante ou paciente (IDP)** em um estudo, A metanálise IDP é reconhecida como a metodologia padrão ouro para a realização de uma metanálise. Ela permite que os pesquisadores: (I) abordem questões não levantadas nas publicações originais (por exemplo, determinar preditores de resultados em um estudo que tenha um objetivo alternativo); (II) usem definições comuns, codificação e pontos de corte; (III) expliquem a variabilidade nos tempos de acompanhamento clínico; e (IV) aumentem o poder estatístico na identificação de características dos participantes.[1,2]

No entanto, devido à baixa taxa de resposta, não conseguimos obter IDP para um número significativo de estudos relevantes. Portanto, realizamos uma metanálise tradicional de estatísticas em nível de estudo, além da metanálise de IDP. Em um estudo sobre dislipidemia, por exemplo, os dados individuais dos participantes poderiam ser os perfis lipídicos pré e pós-tratamento; o parâmetro a ser avaliado no tratamento e as características clínicas como idade e sexo para cada paciente em cada estudo, são dados que devem constar nos estudos de IDP (Quadro 4-1).[1,2] As revisões do IDP oferecem benefícios relacionados à qualidade dos dados, sendo assim o uso de metanálise de IDP nessa modalidade é o padrão ouro, pois seria a análise estatística de dados brutos individual de todas as variáveis dependentes e independentes de cada participante de pesquisas que fossem incluídas na revisão sistemática. Características estatísticas e clínicas podem aumentar o poder para detectar efeitos diferenciais de tratamento em indivíduos em ensaios clínicos randomizados e permitir o ajuste por fatores de confusão em estudos observacionais.[1-3]

Quadro 4-1. Exemplos de Dados de Participantes Individuais de Ensaios de Dislipidemia que Avaliam o Efeito do Tratamento *versus* Placebo no LDL-C[1,2]

Identificação do estudo	Identificação do paciente	Idade (anos)	Sexo (1 = masculino, 0 = feminino)	Grupo de tratamento (1 = tratamento, 0 = controle)	LDL-C mg/dL Pré-tto	LDL-C mg/dL Pós-tto
1	1	38	0	1	160	132
1	2	42	1	0	99	97

tto: tratamento.

Os passos cruciais para a realização de uma metanálise de IDP são decidir a quantidade de dados individuais dos participantes necessários e obter os dados individuais dos participantes. Com isto, podemos adotar uma abordagem de revisão sistemática pela qual todos os estudos relevantes publicados e não publicados são identificados por meio de uma pesquisa transparente e sistemática, e então os autores do estudo são contatados para fornecer os dados individuais dos participantes. Os autores podem estar mais dispostos a concordar se as razões clínicas e metodológicas para requerer os dados individuais dos participantes são claramente delineadas, e também a ajudar a fazer atualizações regulares sobre os resultados da metanálise de dados de participantes individuais e fornecer o incentivo da autoria conjunta em publicações subsequentes.[2]

Os dados agregados se relacionam com informação média ou estimada sobre todos os indivíduos em um estudo, como o efeito médio do tratamento sobre o LDL-C, a idade média ou a proporção de participantes que são do sexo masculino. Esses dados agregados são derivados dos próprios dados individuais do participante. Tal como acontece com qualquer metanálise, uma metanálise de IDP visa resumir a evidência sobre uma questão clínica em particular de vários estudos relacionados e verificar, por exemplo, se um tratamento é eficaz. A estatística utilizada na metanálise de IDP é crucial e deve preservar o agrupamento de pacientes nos estudos; é inadequado simplesmente analisar dados individuais do participante como se todos fizessem parte de um único estudo.[2,4,5]

Os grupos podem ser mantidos durante a análise usando uma abordagem de **um ou de dois passos**. Na abordagem de “dois

passos", os dados individuais dos participantes são primeiro analisados em cada estudo separado de forma independente usando um método estatístico apropriado para o tipo de dado que está sendo analisado. Por exemplo, um modelo de regressão linear pode ser ajustado para respostas contínuas, como pressão arterial, e no modelo de regressão de Cox pode ser aplicada para tempo até o evento, para dados como a mortalidade, internação, síndromes coronarianas e acidente vascular cerebral. Esta etapa produz dados agregados para cada estudo como uma estimativa média do efeito do tratamento e de seu erro-padrão. Esses dados são então sintetizados em um "segundo passo" usando um modelo adequado para a metanálise de dados agregados, como em estudos em que se avalia o inverso da variância, onde se podem assumir os efeitos fixos ou aleatórios (tratamento) em todos os estudos.[2,4,5]

Em uma abordagem em um "único passo", os IDP de todos os estudos são modelados simultaneamente e são responsáveis pelo agrupamento de participantes em estudos. Essa abordagem novamente requer um modelo específico para o tipo de dados que está sendo sintetizado, juntamente com especificações apropriadas na metanálise (p. ex., de efeitos aleatórios em todos os estudos).[2,4,5] (Fig. 4-1)

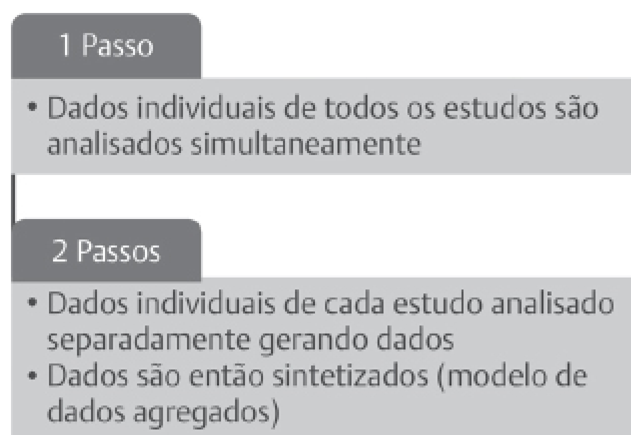


Fig. 4-1. Diferenças entre a metanálise de dados individuais em um e dois passos.

MODELOS ESTATÍSTICOS

Detalhes estatísticos estão disponíveis em artigos sobre as análises e o mérito de métodos de metanálise de dados individuais de um passo ou de dois passos, sendo que as duas abordagens mostraram resultados muito semelhantes, particularmente quando a metanálise visa estimar um único efeito de tratamento de interesse. As metanálises de IDP exigem convenientemente apenas um modelo único, mas isso pode aumentar a complexidade para os não estatísticos e requer uma separação cuidadosa dentro do estudo e entre a variabilidade do estudo.[6-14]

As metanálises de IDP de "dois passos" são claramente mais laboriosas, mas na segunda etapa permitem o uso de técnicas tradicionais bem conhecidas, como as utilizadas pela Cochrane Collaboration (p. ex., variância inversa, efeito fixo ou abordagem de efeitos aleatórios, ou o método de Mantel-Haenszel). Importante, é que tanto as abordagens de um passo como a de dois passos produzem resultados que podem informar evidências científicas. Por exemplo, uma estimativa agregada do efeito do tratamento em todos os estudos e como o efeito do tratamento é modificado pelas características do nível de estudo (p. ex., dose de tratamento ou localização do estudo) e características do nível do paciente (p. ex., idade ou estágio da doença).[15,16]

Em uma "metanálise de dados agregados", os pesquisadores precisam replicar a abordagem de dois passos para a metanálise de dados individuais dos participantes. Por exemplo, a estimativa do efeito do tratamento e de seu erro-padrão é procurada em cada estudo para síntese. Se os dados agregados necessários puderem ser obtidos, um dado de participação individual na metanálise e uma metanálise de dados agregados serão equivalentes, desde que outros fatores sejam iguais (p. ex., número de pacientes, duração do seguimento, e assim por diante).[16-18] Os pesquisadores devem observar que os dados agregados necessários para extrair dependerão das questões clínicas e dos resultados de interesse e das medidas estatísticas mais apropriadas para avaliá-los. Assim, a ligação entre clínicos e estatísticos é crucial para identificar os dados agregados necessários. Também um exercício de escopo ou

evidência de uma revisão semelhante anterior pode ajudar a determinar se tais dados agregados são obtidos e confiáveis.[2]

As análises estatísticas podem ser realizadas usando-se um pacote abrangente de *software* de metanálise (Bio Stat, Inc.) e o Stata version 14 (Stata Corporation, College Station, TX, United States of America).[19]

VANTAGENS DA UTILIZAÇÃO DE UMA ABORDAGEM IDP

As vantagens da utilização de uma abordagem IDP, em vez de dados agregados, para revisão sistemática e metanálise de ensaios clínicos randomizados são apresentadas no Quadro 4-2.[20-24]

Quadro 4-2. Vantagens da Abordagem IDP

Aspecto da revisão sistemática/metanálise	Vantagens
Inclusão experimental	Poder perguntar ao grupo colaborativo do IDP (de avaliadores e outros especialistas no campo clínico) para suplementar a lista de ensaios identificados Esclarecer a elegibilidade do estudo com os investigadores do estudo
Qualidade dos dados	Incluir testes não publicados ou não relatados na íntegra Incluir dados não reportados, mais resultados por avaliação e mais informações completas sobre esses resultados e dados sobre os participantes excluídos das análises de ensaios Verificar a integridade dos IDP de teste e resolver as dúvidas do teste

	com os pesquisadores Estabelecer definições padronizadas de resultados em testes ou traduzir diferentes definições para uma escala comum Definir as classificações padronizadas das características dos participantes ou sua doença/condição ou traduzir diferentes definições para uma escala comum Atualizar o acompanhamento dos resultados de tempo até o evento além do relatado
Risco de viés	Esclarecer o desenho do estudo, conduta e métodos de análise com o julgamento dos pesquisadores Verificar o risco de viés do IDP do teste e obter dados adicionais quando necessário
Análises	Analisar todos os resultados importantes e determinar a validade dos pressupostos de análise com IDP. Por exemplo, proporcionalidade de danos para um modelo de Cox e derivação de medidas de efeito direto do IDP Usar uma unidade consistente de análise para cada tentativa e aplicar um método consistente de análise para cada ensaio Realizar uma análise mais detalhada dos resultados de tempo até o evento, por exemplo gerando curvas de Kaplan Meier Alcançar maior poder para avaliar as

	interações entre os efeitos das intervenções e as características do participante ou doença/condição Realizar mais análises, como a modelagem multinível, para explorar associações entre os efeitos da intervenção e as características do paciente Usar modelos não padronizados ou medidas de efeito, e também utilizar os IDP para questões clínicas secundárias, por exemplo para explorar história de doença, fatores prognósticos ou desfechos substitutos
Interpretação	Discutir as implicações para a prática clínica e a pesquisa com um grupo multidisciplinar de colaboradores, incluindo os investigadores do estudo que forneceram os dados

Critérios claros de elegibilidade são necessários para determinar por que determinados estudos foram incluídos em uma metanálise de IDP e, portanto, se dão uma visão imparcial ou representativa da evidência. Os vieses que podem afetar revisões sistemáticas e metanálises de IDP e dados agregados (DA), assim como as medidas tomadas para investigar e/ou minimizar, são apresentados no Quadro 4-3.[20-24]

Quadro 4-3. Vieses que Podem Afetar Revisões Sistemáticas e Metanálises de IDP

Tipo de viés	Etapas que são tomadas para investigar e minimizar o viés			
		Usual com ambos as abordagens de IDP e DA	**Habitualmente com a abordagem de IDP, mas pode ser possível com abordagem de DA***	**Somente com IDP**
Viés de seleção de estudo: Diferenças sistemáticas entre os resultados dos estudos que são e não são selecionados para inclusão	Critérios de elegibilidade definidos prospectivamente	X		
	Esclareça a elegibilidade com o protocolo de ensaio ou do ensaio		X	
Viés de publicação: Diferenças sistemáticas entre resultados de estudos que são e não são publicados	Incluir todos os testes elegíveis, independentemente do *status* de publicação		X	
Viés de disponibilidade de dados: Diferença sistemática entre os resultados de ensaios para os quais os dados estiveram ou não estiveram disponíveis	Incluir dados para todos os testes qualificados Investigar/discutir o impacto de ensaios para os quais os dados não foram acessíveis		X	
Viés de seleção de participantes: Diferenças sistemáticas entre grupos de comparação por características dos participantes que podem levar a diferenças no prognóstico e/ou capacidade de resposta ao tratamento (evitadas por alocação aleatória e ocultação de alocação)	Esclarecer os métodos de randomização, ou seja, sequência, geração e ocultação de alocação com protocolo de teste ou ensaio Excluir ensaios “não randomizados”		X	
	Verificar padrões de alocação incomuns ou distribuições de características dos participantes Excluir tentativas com alocação imprópria Excluir participantes não randomizados do estudo IDP			X

Desempenho e viés de detecção: Diferenças sistemáticas entre grupos de comparação no cuidado recebido ou fornecido ou em como os resultados são determinados (evitadas por cegar os participantes do estudo, prestadores de cuidados e avaliadores de resultados para o tratamento alocado. Observe que isso não é possível para todas as intervenções – por exemplo, cirurgia – e é menos importante para resultados objetivos – por exemplo, mortalidade	Obter informações mais completas sobre cegamento e avaliação de resultados do avaliador e/ou protocolo		X	
Viés de atrito: Diferenças sistemáticas entre grupos de comparação no abandono ou exclusão de participantes (evitadas pela manutenção de todos os participantes na análise do estudo e teste)	Incluir dados sobre todos os participantes randomizados, independentemente se eles foram incluídos em análises de testes Análise de todos os ensaios de acordo com a intervenção alocada ("intenção de tratar")		X	
	Verificar se há participantes "ausentes" e padrões incomuns de abandono ou exclusão Pré-especificar quaisquer exclusões razoáveis de participantes e aplicar consistentemente por meio de ensaios			X
Relatório de resultados ou viés de disponibilidade: Diferenças sistemáticas entre os resultados relatados/disponíveis e não relatados/indisponíveis (evitadas por meio da disponibilização de resultados para todos os resultados do estudo)	Verifique quais resultados foram coletados em um estudo com protocolo e/ou com o trialista Incluir dados para todos os resultados relevantes		X	

*Somente possível com revisões de DA se dados detalhados e de alta qualidade e/ou outras informações puderem ser obtidos de protocolos de estudos, relatórios de testes, ou diretamente do estudo ou com os investigadores.

As perguntas a serem feitas ao se avaliar criticamente uma metanálise de IDP são estas abaixo:[20-24]

1. **A metanálise de IDP faz parte de uma revisão sistemática?**

a. Tem uma pergunta clara de pesquisa qualificada por critérios explícitos de elegibilidade?
b. Possui uma estratégia de busca sistemática e abrangente para identificar ensaios?
c. Tem uma abordagem consistente para a coleta de dados?
d. Avalia a “qualidade” ou o risco de viés dos estudos incluídos?
e. Todos os métodos são pré-especificados em um protocolo?
f. O protocolo foi registrado ou de outra forma disponibilizado?

2. **Todos os estudos elegíveis foram identificados?**
    a. Foram totalmente identificados os ensaios publicados?
    b. Os ensaios publicados na literatura cinzenta foram identificados?
    c. Os ensaios não publicados foram identificados?
3. **O IDP foi obtido para a maioria dos ensaios?**
    a. IDP obtido para uma grande proporção dos ensaios elegíveis?
    b. Foi feita uma avaliação do impacto potencial dos ensaios em falta?
    c. As razões para não obter IDP foram fornecidas?
4. **A integridade do IDP foi verificada?**
    a. Os dados foram verificados em busca de itens ausentes, inválidos, fora de alcance ou inconsistentes?
    b. Houve alguma discrepância com o relatório do teste (se disponível)?
    c. Algum problema foi questionado e, se possível, resolvido?
5. **As análises foram pré-especificadas em detalhe?**
    a. Os métodos de análise detalhados foram incluídos em um protocolo ou plano de análise?
    b. Foram incluídos os resultados e métodos para analisar os efeitos das intervenções, quantificando e explicando a heterogeneidade e avaliando o risco de viés incluído?
6. **O risco de viés dos estudos incluídos foi avaliado?**
    a. A randomização, a ocultação da alocação e o cegamento foram avaliados?
    b. Os IDP foram verificados para garantir que todos (ou a maioria) dos participantes randomizados foram incluídos?

c. Todos os resultados relevantes foram incluídos?
d. A qualidade dos dados do tempo até o evento foi verificada?

7. **Os métodos de análise eram apropriados?**
   a. Os métodos de avaliação dos efeitos globais das intervenções foram apropriados?
      a1. Os pesquisadores estratificaram ou explicaram o agrupamento de participantes nos ensaios usando uma abordagem de uma ou duas etapas para metanálise?
      a2. A escolha da análise de uma ou duas etapas foi especificada com antecedência e/ou os resultados para ambas as abordagens fornecidas?
   b. Os métodos para avaliar se os efeitos das intervenções variaram de acordo com as características do ensaio eram apropriados?
      b1. Os pesquisadores compararam os efeitos do tratamento entre subgrupos de ensaios ou fizeram uma metarregressão para avaliar se o efeito global do tratamento varia na testagem de suas características?
   c. Os métodos de avaliar se os efeitos das intervenções variam de acordo com o participante apresentam características apropriadas?
      c1. Os pesquisadores estimaram uma interação separadamente para cada tentativa e combinaram estes por meio de ensaios em um efeito fixo de dois estágios ou metanálise de efeitos aleatórios?
      c1. Os pesquisadores incorporaram uma ou mais covariáveis de um tratamento por participante em termos de interação em um modelo de regressão ao mesmo tempo em que cada participante separava essa interação individual no nível do participante de qualquer interação em nível de teste?
   d. Se não houve evidência de um efeito diferencial por prova ou característica do participante, foi colocada ênfase no resultado global?
   e. As análises exploratórias foram destacadas como tal?

8. **Algum relatório dos resultados adere aos Itens de Relatório Preferidos para uma revisão sistemática e metanálise de IDP (a declaração PRISMA-IDP)?[25]**

Um estudo recente de metanálise IDP mostrou que a estratégia de tratamento para reduzir a pressão arterial com base no risco cardiovascular previsto é mais eficaz do que uma baseada apenas nos níveis de pressão arterial em uma série de limiares. Estes resultados apoiam a avaliação de risco cardiovascular para orientar a tomada de decisão sobre o tratamento da pressão arterial em indivíduos de risco moderado a alto, particularmente na prevenção primária.[19]

CONCLUSÃO

A decisão de realizar uma metanálise de IDP deve ser conduzida por questões clínicas de interesse e, se uma metanálise de dados agregados publicada pode respondê-las de forma cientificamente confiável, uma metanálise de IDP pode oferecer vantagens consideráveis, tanto estatística quanto clinicamente.

REFERÊNCIAS BIBLIOGRÁFICAS

1. Simmonds MC, Higgins JP, Stewart LA, Tierney JF, Clarke MJ, Thompson SG. Meta-analysis of individual patient data from randomized trials: a review of methods used in practice. Clin Trials. 2005;2(3):209-217.
2. Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. BMJ. 2010;340:c221.
3. Stewart LA, Tierney JF. To IPD or not to IPD? Advantages and disadvantages of systematic reviews using individual patient data. Eval Health Prof. 2002;25(1):76-97.
4. Burgess S, White IR, Resche-Rigon M, Wood AM. Combining multiple imputation and meta-analysis with individual participant data. Stat Med. 2013;32(26):4499-4514.
5. Debray TP, Moons KG, Ahmed I, Koffijberg H, Riley RD. A framework for developing, implementing, and evaluating clinical

prediction models in an individual patient data meta-analysis. Stat Med. 2013;32(18):3158-3180.

6. Turner RM, Omar RZ, Yang M, Goldstein H, Thompson SG. A multilevel model framework for meta-analysis of clinical trials with binary outcomes. Stat Med. 2000;19:3417-32.
7. Riley RD, Dodd SR, Craig JV, Thompson JR, Williamson PR. Meta-analysis of diagnostic test studies using individual patient data and aggregate data. Stat Med. 2008;27:6111-36
8. Riley RD, Lambert PC, Staessen JA, Wang J, Gueyffier F, Thijs L, et al. Meta-analysis of continuous outcomes combining individual patient data and aggregate data. Stat Med. 2008;27:1870-93.
9. Jones AP, Riley RD, Williamson PR, Whitehead A. Meta-analysis of individual patient data versus aggregate data from longitudinal clinical trials. Clin Trials. 2009;6:16-27.
10. Smith CT, Williamson PR, Marson AG. Investigating heterogeneity in an individual patient data meta-analysis of time to event outcomes. Stat Med. 2005;24(9):1307-19.
11. Whitehead A, Omar RZ, Higgins JP, Savaluny E, Turner RM, Thompson SG. Meta-analysis of ordinal outcomes using individual patient data. Stat Med. 2001;20:2243-60.
12. Higgins JP, Whitehead A, Turner RM, Omar RZ, Thompson SG. Meta-analysis of continuous outcome data from individual patients. Stat Med. 2001;20:2219-41.
13. Olkin I, Sampson A. Comparison of meta-analysis versus analysis of variance of individual patient data. Biometrics. 1998;54:317-22.
14. Mathew T, Nordstrom K. On the equivalence of meta-analysis using literature and using individual patient data. Biometrics. 1999;55:1221-3.
15. Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions version 5.0.1 [updated September 2008]. The Cochrane Collaboration, 2008.
16. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. J Natl Cancer Inst. 1959;22:719-48.

17. Ahmed I, Sutton AJ, Riley RD. Assessment of publication bias, selection bias, and unavailable data in meta-analyses using individual participant data: a database survey. BMJ. 2012;344:d7762.
18. Vale CL, Rydzewska LHM, Rovers MM, Emberson JR, Gueyffier F, Stewart LA, et al. Cochrane IPD Meta-analysis Methods Group. Uptake of systematic reviews and meta-analyses based on individual participant data in clinical practice guidelines: descriptive study. BMJ. 2015;350:h1088.
19. Karmali KN, Lloyd-Jones DM, van der Leeuw J, Goff DC Jr, Yusuf S, Zanchetti A, et al. Blood pressure-lowering treatment strategies based on cardiovascular risk versus blood pressure: A meta-analysis of individual participant data. PLoS Med. 2018;15(3):e1002538.
20. Tierney JF, Vale C, Riley R, Smith CT, Stewart L, Clarke M, et al. Individual Participant Data (IPD) Meta-analyses of Randomized Controlled Trials: Guidance on Their Use. PLoS Med 2015;12(7):e1001855.
21. Cooper YA, Pianka ST, Alotaibi NM, Babayan D, Salavati B, Weil AG, et. al. Repetitive transcranial magnetic stimulation for the treatment of drug-resistant epilepsy: A systematic review and individual participant data meta-analysis of real-world evidence. Epilepsia Open. 2017;3(1):55-65.
22. Maxwell L, Nandi A, Benedetti A, Devries K, Wagman J, García-Moreno C. Intimate partner violence and pregnancy spacing: results from a meta-analysis of individual participant time-to-event data from 29 low-and-middle-income countries. BMJ Glob Health. 2018;3(1):e000304.
23. Purgato M, Gross AL, Betancourt T, Bolton P, Bonetto C, Gastaldon C, et. al. Focused psychosocial interventions for children in low-resource humanitarian settings: a systematic review and individual participant data meta-analysis. Lancet Glob Health. 2018;6(4):e390-e400.
24. Chen B, Benedetti A. Quantifying heterogeneity in individual participant data meta-analysis with binary outcomes. Syst Rev. 2017:6;6(1):243.

25. Stewart LA, Clarke M, Rovers M, Riley RD, Simmonds M, Stewart G, et al. Preferred Reporting Items for a Systematic Review and Meta-analysis of Individual Participant Data: The PRISMA-IPD Statement. JAMA. 2015;313(16):1657-65.

COMPREENDENDO O GRADE

CAPÍTULO 5

RESUMO

A saúde baseada em evidências refere-se ao uso criterioso, cuidadoso e preciso do conhecimento científico existente oriundo de pesquisas clínicas utilizando-se de metodologias específicas que garantam uma solidez e clareza nas informações a serem aplicadas na tomada de decisão clínica, reduzindo assim as incertezas no julgamento clínico frente ao paciente. O objetivo deste capítulo é descrever a metodologia do PICO e a qualidade dos estudos do sistema GRADE.

Palavras-chave: GRADE; metodologia; saúde baseada em evidências.

INTRODUÇÃO

A saúde baseada em evidências (SBE) refere-se ao uso criterioso, cuidadoso e preciso do conhecimento científico existente oriundo de pesquisas clínicas utilizando-se de metodologias específicas que garantam uma solidez e clareza nas informações a serem aplicadas na tomada de decisão clínica, reduzindo assim as incertezas na tomada de decisão junto ao paciente.[1-3]

As recomendações sobre um comparador serão aplicáveis, por exemplo, quando as consequências desejáveis superam as indesejáveis. Isto favorece a aplicação de um conhecimento, técnica

ou terapia, ou do oposto, com a recomendação contrária ao uso, quando os benefícios não suplantam as possíveis complicações ou se igualam, facilitando a aplicação correta para o paciente correto no momento ou fase da doença que tragam resultados positivos em desfechos relevantes.[4,5]

A Avaliação, Desenvolvimento e Avaliação da Classificação de Recomendações (GRADE) desenvolveram uma abordagem padronizada para classificar o valor da evidência e a força das recomendações. Com a abordagem GRADE, os resultados foram classificados como críticos e essenciais, ou não críticos, e apresentados nos perfis de evidências.[5] Como consequência, muitas sociedades internacionais usaram-na como padrão no desenvolvimento de diretrizes (Quadro 5-1).[2]

PICO

O primeiro passo crucial é formular uma pergunta bem focalizada e a identificação dos recursos apropriados, pois a busca por evidências relevantes pode ser demorada. Uma pergunta deve ter quatro componentes: paciente/população, intervenção (diagnóstica ou terapêutica), alternativa à intervenção (comparação) e os resultados de interesse (desfechos) (Quadro 5-2).

Quadro 5-1. Qualidade da Evidência

Itens	Critério	Aplicação
Fatores que diminuem a qualidade da evidência		
1. Limitações do estudo (risco de viés)	Resultado da avaliação metodológica de cada delineamento	Diminuir 1 ponto se o risco de viés for considerado sério ou 2 pontos se for muito sério
2. Inconsistência dos resultados (heterogeneidade)	No caso de desfechos inconsistentes,	Reduzir 1 ponto caso a

	avaliar semelhança das estimativas, sobreposição dos intervalos de confiança e resultados dos testes de heterogeneidade e do I^2	inconsistência seja importante
3. Evidência indireta	Avaliar se existem diferenças na população, intervenção, comparação ou desfechos entre os estudos incluídos e a pergunta de interesse da revisão	Rebaixar 1 ponto se a evidência indireta for séria ou 2 pontos se for muito séria
4. Imprecisão	Avaliar a amplitude do intervalo de confiança, ou se o número de eventos e o tamanho da amostra são pequenos	Reduzir 1 ou 2 pontos se houver imprecisão
5. Viés de publicação	Avaliar se há possibilidade de estudos não terem sido publicados, bem como a influência dos financiamentos da pesquisa	Rebaixar 1 ponto caso haja suspeita de viés de publicação
Fatores que aumentam a qualidade da evidência (aplicável aos estudos observacionais)		

1. Grande magnitude de efeito	A observação de grande efeito aumenta a confiança na evidência encontrada	Elevar a qualidade em 1 ponto (se RR ≥ 2 ou ≤ 0,5) ou 2 pontos (se RR ≥ 5 ou ≤ 0,2)
2. Gradiente dose-resposta	A observação de alteração do efeito conforme a exposição se modifica auxilia na definição da causalidade	Aumentar a qualidade em 1 ponto se houver gradiente dose-resposta
3. Confundidores ou vieses reduziriam o efeito encontrado	A presença de confundidores (que estariam indo na direção oposta ao efeito) não impede que o resultado favorável à intervenção seja encontrado	Aumentar a qualidade em 1 ponto se os confundidores existentes diminuíram o efeito observado

Os problemas potenciais das perguntas podem ser as falhas da avaliação de todas as alternativas relevantes (p. ex., em diretrizes onde as opções de tratamento variam em diferentes países). Dois outros erros intimamente relacionados são a exclusão de resultados relevantes e a ênfase excessiva nos resultados substitutos com importância questionável.[3] O quadro PICO facilita o processo de busca, identificando os fundamentos para uma estratégia de busca eficaz.[6,8]

As magnitudes de efeito dos resultados em termos relativos (p. ex., risco relativo, taxa de risco e razão de chances) são apresentadas na maioria dos estudos originais e revisões sistemáticas. No entanto, a diferença entre as consequências desejáveis e indesejáveis de uma recomendação requer o conhecimento dos efeitos em uma população específica. O mérito

do quadro PICO é a sua transparência sobre a lógica por trás da escolha, e o equilíbrio entre resultados desejáveis e indesejáveis determina a direção e influencia a força de uma recomendação. Esse equilíbrio depende da magnitude dos esperados efeitos desejáveis e indesejáveis e como os pacientes valorizam resultados.[9]

Quadro 5-2. Elementos do PICO

Questão	**PICO**	**Etapas para a formulação da questão clínica usando o formato PICO**
Desfecho	Estimar o efeito para cada desfecho	Enquadrar a questão ■ Escreva as informações necessárias na forma de uma pergunta ■ A questão deve incluir os elementos do PICO P = Problema, paciente ou população I = Intervenção/indicador C = Comparação O = Resultado de interesse
Taxa de qualidade	Alta Moderada Baixa Muito baixa	Planejar uma estratégia de pesquisa identificando os

		elementos significativos da pergunta, e traduzir termos de linguagem para descritores de assunto ou termos MeSH Para iniciar a pesquisa e manter os resultados da pesquisa inicial amplos, comece com o P e o I
Recomendações	Forte: as vantagens de uma determinada conduta claramente suplantam as desvantagens da mesma, ou então as desvantagens claramente suplantam as vantagens. Fraca: há certo grau de incerteza sobre a relação entre vantagens e desvantagens de uma dada conduta.	Após os resultados da pesquisa inicial, faça uma pesquisa restrita com os termos da comparação, resultado, fatores de tempo ou tipo de estudo Veja resumos e textos completos de artigos para analisar a comparação e o resultado Faça o *download* da planilha booleana para planejar a pesquisa no banco de dados

Embora valha a pena especificar os resultados críticos e essenciais antes de começar a revisão das evidências, esta revisão pode influenciar os julgamentos e a importância dos resultados.[1,2]

Em duas situações, as análises da revisão de evidências podem modificar a seleção de resultados relevantes ou sua importância relativa. Um benefício potencial de um resultado inicialmente julgado como fundamental poderia deixar de sê-lo em uma revisão e, frente a dados outros benefícios estabelecidos, a intervenção poderia ser julgada viável na ausência de um benefício demonstrado. Por exemplo, na triagem de uma doença para a qual se está usando a droga A, é provável que se considere o impacto da intervenção em todas as causas de mortalidade como crítica.[1,2]

No entanto, o resumo das evidências estabelece uma redução essencial na mortalidade específica por causa de câncer de próstata, mas falha em estabelecer uma redução definitiva da mortalidade por todas as causas. Portanto, a redução na mortalidade específica por causa pode ser julgada suficientemente convincente, mesmo na ausência de uma redução da mortalidade por todas as causas. A mortalidade por todas as causas então se torna menos relevante e deixa de ser um resultado crítico.[6]

Por outro lado, qualquer nova intervenção poderia estar associada com efeitos adversos que não estavam inicialmente presentes. Inesperada e significativa toxicidade poderia surgir em drogas aprovadas, e a não descoberta da toxicidade deve ser considerada como um efeito adverso, consequência de qualquer novo medicamento. Portanto, a toxicidade torna-se crítica somente quando evidências suficientes de sua existência emergem.[6]

AVALIAÇÃO DA QUALIDADE DOS ESTUDOS

Após a busca do PICO, o GRADE faz uma classificação preliminar dos resultados e garante que as diretrizes abordem as questões e a qualidade da evidência de maneira sistemática.[6,9]

Os desenvolvedores de diretrizes consideram determinados recursos em ensaios controlados, como randomização, ocultação de alocação e cegamento. Análise de intenção de tratamento pode ser usada para gerar uma classificação. Em estudos observacionais, a medição apropriada da exposição e o controle adequado dos fatores de confusão devem ser avaliados. Em ambos os ensaios controlados e observacionais, os desenvolvedores devem levar em

conta a perda de acompanhamento, bem como outros aspectos da concepção, condução e análise da investigação que podem influenciar o risco de viés.[7]

Cada uma das abordagens destina-se a usar os dados disponíveis para verificar a probabilidade de um viés de publicação. Um registro prospectivo de todos os ensaios clínicos randomizados permite aos autores das revisões avaliarem quando os ensaios relevantes foram realizados porque é possível pedir aos investigadores responsáveis que forneçam as informações dos dados do estudo. O registro obrigatório destes ensaios clínicos representa um dos métodos confiáveis para evitar um viés de publicação.[7,8]

Consequentemente, uma pesquisa de registros de ensaios clínicos deve ser considerada ao se avaliar o risco de viés de publicação.[8] A categorização final de evidências baseadas no escore GRADE não se relaciona necessariamente com a qualidade metodológica global de qualquer ensaio clínico controlado randomizado ou revisão. De certa forma, se relaciona com a qualidade da evidência em um resultado específico em nossa população designada de interesse.[1,2]

Na literatura, a qualidade comumente se refere a um julgamento sobre a validade interna de um estudo individual. Embora outros significados tenham sido atribuídos à palavra "qualidade" (tipicamente risco de viés), este significado deve corresponder mais de perto ao padrão e compreensão não técnica da qualidade.[9]

O GRADE não se refere a estudos individuais, mas a um corpo de evidências. Um corpo de evidências (p. ex., alguns ensaios bem desenhados e executados) pode estar associado a um baixo risco de viés, mas as confianças estimativas podem ser comprometidas por alguns outros fatores (imprecisão, inconsistência, indireção e viés de publicação). Tem também fatores particularmente relevantes para estudos observacionais, pois podemos avaliar a qualidade, incluindo a magnitude do efeito do tratamento e a presença de um gradiente dose-resposta. Estudos bem conduzidos podem fazer parte de um conjunto de evidências classificadas como de baixa

qualidade porque eles fornecem apenas evidências indiretas ou imprecisas para a questão de interesse.[9]

Classificando na abordagem GRADE cada um dos resultados de interesse como importante, mas não crítico, ou não relevante para fazer recomendações, especialistas, clínicos e pacientes diferem no valor de determinados resultados. Por exemplo, resultados como mortalidade ou qualidade da vida podem ser considerados sintomas críticos, segundo um profissional de saúde, ou valioso, mas não crítico, ou não relevante, mas talvez informativo para fazer uma recomendação.[3]

No sistema GRADE, o desenho do estudo continua a ser crítico ao julgar a qualidade da evidência. Portanto, no sistema GRADE, um corpo de evidências obtidas a partir de ensaios clínicos é inicialmente classificado como de alta qualidade e aquele obtido a partir de observações de estudos, como de baixa qualidade.[1,2]

Um bem projetado e executado ensaio randomizado ou estudo observacional fornece uma qualidade diferente em termos de evidência do um que foi malconduzido. Assim sendo, confiar apenas no desenho do estudo tem limitações aparentes. O GRADE fornece critérios adicionais de qualidade que servem para superar essa lacuna. Portanto, cinco fatores podem reduzir a qualidade de evidências para cada desenho de estudo, e três podem aumentá-lo.[1,2]

RISCO DE VIÉS (LIMITAÇÕES NO DESENHO DO ESTUDO E/OU EXECUÇÃO)

A qualidade diminui quando os estudos sofrem de significativa limitação metodológica, o que pode influenciar suas estimativas do efeito do tratamento como a falta de ocultação de alocação, falta de cegamento (resultados subjetivos e avaliação altamente suscetível a vieses), falta de contabilizar uma proporção significativa de pacientes que iniciaram o estudo (perda considerável no seguimento ou o resultado não medido em uma proporção significativa de pacientes), falha em aderir ao princípio da intenção de tratar durante a análise, pareamento para obter benefícios, ou relatar seletivamente os resultados que mostram um aparente efeito de

tratamento e falha em relatar resultados que não mostram nenhum efeito evidente.[1,2]

Limitações metodológicas implicam uma maior propensão a vieses, diminuindo, assim, a confiança na estimativa de efeito. Na presença de limitações, elas podem ser classificadas em graves ou muito graves, reduzindo respectivamente em um ou dois níveis sua graduação de qualidade de acordo com sua qualitativa propensão de enviesar os resultados. A seguir seguem listados alguns dos principais limitantes metodológicos de ensaios clínicos e de estudos observacionais.[1,2]

Potenciais Limitações em Ensaios Clínicos

- Sigilo de alocação ou randomização inadequada:
  - Estudos não randomizados (quase experimentos que, apesar de iniciarem com pontuação máxima devido ao delineamento, são obrigatoriamente penalizados neste critério). Métodos de randomização não aleatórios (por exemplo, data de nascimento, ordem de chegada). Alocação não sigilosa, como listas abertas com sequência de randomização que permita ao investigador inferir sobre o grupo de intervenção destinado ao participante antes do mesmo ser incluído no estudo.[1,2]
- Mascaramento inadequado ou ausente:
  - Mascaramento inadequado ou ausente pode implicar em viés, especialmente quando referente a avaliadores de desfechos. Por exemplo, ausência de mascaramento dos avaliadores em desfechos avaliador-dependentes como qualidade de vida e escalas psicrométricas.[1,2]
- Descrição inadequada de seguimento de todos os pacientes inicialmente randomizados:
  - Perda de seguimento (em especial quando superior a 10-20%). Ausência de relato adequado do motivo e dos grupos das perdas, quando presentes. Análise não realizada por intenção de tratamento, quando indicada.[1,2]
- Relato seletivo de desfechos:

- Relato dos desfechos de acordo com os resultados encontrados; por exemplo, hipóteses não definidas *a priori*.

- Fenômeno *carry-over* em estudos cruzados (*cross-over*):
  - Intervenção realizada na primeira fase de um estudo cruzado interfere na segunda fase, como, por exemplo, medidas educacionais ou uso de antibióticos que podem alterar a flora bacteriana.[1,2]
- Interrupção precoce de ensaio clínico devido a benefício, em especial na ausência de regras predefinidas para interrupção dos estudos.
- Viés de seleção em ensaios clínicos com randomização de *clusters*:
  - Grupos de clusters não comparáveis. Por exemplo, randomização de UTIs em vez de pacientes, com os grupos resultantes não adequadamente comparáveis entre si.[1,2]

POTENCIAIS LIMITAÇÕES EM ESTUDOS OBSERVACIONAIS

- Vieses de seleção.
  - Pareamento inadequado em estudos de caso-controle. Seleção de indivíduos expostos e não expostos procedentes de diferentes populações em estudos de coorte. Seleção inadequada de controles em estudos de caso-controle.[1,2]
- Vieses de aferição.
  - Diferenças nas medidas da exposição nos diferentes grupos (por exemplo, viés de recordação em estudos de caso-controle). Vigilância diferente de desfechos em expostos e não expostos.[1,2]
- Vieses de confusão.
  - Presença de potenciais confundidores não avaliados. Análise estatística sem adequado ajuste de confundidores.[1,2]
- Seguimento inadequado.

- Perdas de seguimento (em especial superiores a 20%). Período de seguimento pequeno para a associação em avaliação.[1,2]

INCONSISTÊNCIA DOS RESULTADOS

A qualidade das evidências diminui se forem amplamente diferentes as estimativas de efeito do tratamento em estudos individuais (variabilidade ou heterogeneidade dos resultados) e que sugerem diferenças exatas nos efeitos do tratamento. A variabilidade nos resultados individuais do estudo pode dar diferenças clínicas nas populações, nas intervenções e nos resultados medidos ou de limitações metodológicas (problemas com randomização ou término antecipado dos ensaios).[1,2]

É um tópico importante a ser avaliado em metanálises ou em conjuntos de estudos individuais. Neste caso, inconsistência significa que estudos com hipóteses semelhantes possuem estimativas diferentes do que se esperaria simplesmente pelo acaso. Quando há inconsistência importante nos resultados e as razões para essas diferenças não são adequadamente exploradas e explicadas, o efeito encontrado pode não corresponder à realidade, devendo a evidência ter seu nível de qualidade reduzido em um ou dois níveis.[1,2]

Essa variabilidade dos resultados geralmente prova heterogeneidade (diferença) entre os estudos referente a:

- *População:* por exemplo, AAS gera maior benefício na prevenção de eventos cardiovasculares em indivíduos mais velhos.
- *Intervenções:* diferenças no tipo, dose ou via de administração de uma determinada intervenção; por exemplo, medicações de liberação prolongada têm melhores benefícios, pois implicam maior aderência ao tratamento.
- *Desfechos:* critérios diferentes para avaliação de um mesmo desfecho, diminuição do efeito do tratamento com o aumento do tempo de observação; por exemplo, benefício maior da sibutramina na perda de peso nas 8 semanas iniciais de administração.

Em uma metanálise, há testes que auxiliam na identificação da variabilidade entre os estudos. O teste de heterogeneidade detecta a presença de heterogeneidade estatística na amostra (presente quando $p < 0,1$) e no teste de inconsistência (I2), é considerada inconsistência alta, em geral, quando superior a 50%. Esses testes podem auxiliar na análise da presença de diversidade entre os estudos.[1,2] A inconsistência, por vezes, pode ser explicada por análise de subgrupo e metarregressão. Contudo, a presença de inconsistência não explicada ou parcialmente explicada pode prejudicar a confiança na medida de associação agregada encontrada, diminuindo a qualidade da evidência.[1,2]

INDIREÇÃO DE EVIDÊNCIA

O GRADE distingue dois tipos de indireção. A comparação indireta surge quando a recomendação aborda a escolha entre, por exemplo, duas drogas ativas, mas os estudos disponíveis compararam os medicamentos com placebo. Uma comparação indireta fornece evidências de menor qualidade do que uma comparação frente a frente e é comum quando se escolhe entre drogas da mesma classe.[1,2] As evidências apoiando a recomendação também são indiretas quando se trata de estudos nos quais população, intervenção, intervenção alternativa ou resultados de interesse eram diferentes daqueles recomendados.

Corresponde a transposição das evidências disponíveis de recomendações correlatas para a recomendação em questão proposta.[1,2]

Dois tipos principais:

1. Comparação indireta. Não há estudos comparando diretamente a intervenção A *vs.* B; contudo, há estudos comparando intervenção A *vs.* placebo e intervenção B *vs.* placebo. Esses estudos podem ser agregados para comparar as intervenções A e B (como é o caso de *multiple treatment comparions* e *network metanalysis*); contudo, há perda substancial na confiança dos efeitos estimados.

2. População, intervenção, comparadores ou desfechos indiretos, a evidência não corresponde exatamente ao mesmo cenário que se quer transpor.

IMPRECISÃO DOS RESULTADOS

As estimativas do efeito geralmente têm amplos intervalos de confiança, que incluem tanto benefícios essenciais quanto não efeitos essenciais (ou mesmo danos) quando os estudos apontam, relativamente, que poucos eventos ocorrem. Com a incerteza resultante em vigor, a indeterminação dos resultados poderia diminuir a qualidade da evidência. Corresponde ao erro aleatório, gerando imprecisão na estimativa (intervalos de confiança amplos) devido à amostra ser pequena ou ao pequeno número de eventos. A imprecisão dos resultados prejudica a qualidade da evidência pela incerteza na magnitude do benefício ou do risco da intervenção.[1,2]

A decisão de penalizar os estudos devido à imprecisão deve-se a uma série de fatores como tamanho de amostra, raridade do evento etc., devendo ser avaliada qualitativamente para cada situação. Entretanto, em geral, sugere-se imprecisão nos resultados quando: em metanálise, o tamanho cumulativo da amostra é inferior ao tamanho ótimo da informação calculado, o número de eventos totais é pequeno (inferior a 300), ou então, para desfechos contínuos (p. ex., qualidade de vida) a amostra total seja inferior a 400 indivíduos, quando um intervalo de confiança de 95% apresenta ausência de efeito e considerável benefício ou risco (esse geralmente sendo considerado de aumento ou diminuição do risco relativo da ordem de 25%).[1,2]

VIÉS DE PUBLICAÇÃO

A qualidade da evidência será reduzida se os pesquisadores não informarem os estudos nos quais nenhum efeito é apresentado. O risco de viés é maior se apenas alguns pequenos estudos estiverem disponíveis. No entanto, enquanto pequenos estudos relataram benefícios clínica e estatisticamente significativos do tratamento

ativo, os estudos maiores mostraram um efeito estatisticamente não significativo.[1,2]

Estudos com resultados positivos são publicados em forma de artigo, com maior frequência em revistas de maior impacto, geralmente em língua inglesa, e em um período de tempo menor (e por vezes gerando também mais do que uma publicação com a mesma base de dados). Além disso, o relato seletivo de desfechos não é incomum, podendo ser publicados apenas os desfechos cujos resultados se mostraram favoráveis, desconsiderando os demais.[1,2] Como resultante, temos metanálises propensas a apresentar benefício espúrio a favor de determinadas intervenções ou, então, propensas a aumentar artificialmente sua magnitude, dessa forma prejudicando nossas conclusões a respeito da questão da pesquisa. Diversas alternativas são utilizadas para avaliar o viés de publicação, como testes estatísticos (teste de Egger) e análises gráficas (*funnelplot*). Na presença de viés de publicação com potencial impacto nas conclusões dos resultados, a qualidade da evidência deverá ser reduzida.[1,2]

GRANDE MAGNITUDE DE EFEITO

Quando baseado em estudos de modelagem que fornecem estimativas da magnitude do efeito não explicado por um viés, o sistema GRADE define um risco relativo < 0,5 ou > 2 como o efeito significativo e um risco relativo < 0,2 ou > 5 como efeito substancial.[1,2] Quando estudos, mesmo observacionais, com forte rigor metodológico estimam associações consistentes e com magnitude importante, aumenta nossa confiança nos resultados.[1,2] Quanto maior o tamanho do efeito, mais forte se torna a evidência. Importante salientar que estudos que tiveram seus níveis de evidência reduzidos provavelmente possuem validade interna comprometida, sendo desaconselhável aumentar sua qualidade.[1,2] Em geral, podemos adotar o critério apresentado no Quadro 5-3.

Quadro 5-3. Tamanho do Efeito

Tamanho do efeito	Medida do efeito	Qualidade
Grande	RR > 2 ou < 0,5 (baseado em evidências consistentes de pelo menos dois estudos sem possíveis confundidores)	Aumentar um nível
Muito grande	RR > 5 ou < 0,2 (baseado em evidência direta sem importante comprometimento da validade)	Aumentar dois níveis

REDUÇÃO OU AUMENTO DO EFEITO DEMONSTRADO POR CONFUNDIMENTO PLAUSÍVEL

Todos os vieses plausíveis podem subestimar o efeito exato do tratamento, considerando as evidências desses estudos observacionais de baixa qualidade. Uma situação paralela existe quando estudos observacionais não demonstram uma associação, mas todos os vieses aumentaram um efeito de intervenção e levantaram efeitos prejudiciais aparentes.[1,2]

A presença de confundidores que, se ajustados, provavelmente levariam a resultados na direção contrária à observada, aumenta nossa confiança a respeito da direção do efeito encontrado. Se o grupo recebendo uma determinada intervenção (ou exposição) for composto por pacientes mais doentes do que o respectivo grupo controle, e mesmo assim obterem melhores resultados, é esperado que, caso os grupos fossem iguais, o benefício da intervenção seria ainda maior. Nesse caso, o nível da qualidade da evidência, caso

não haja comprometimento importante de sua validade interna, poderá ser aumentado.[1,2]

GRADIENTE DOSE-RESPOSTA

Um gradiente dose-resposta também pode aumentar a qualidade das evidências. Além disso, quanto maior a exposição, maior o risco.[3] Se o grupo recebendo uma determinada intervenção (ou exposição) for composto por pacientes mais doentes do que o respectivo grupo controle, e mesmo assim obterem melhores resultados, é esperado que, caso os grupos fossem iguais, o benefício da intervenção seria ainda maior. Nesse caso, o nível da qualidade da evidência, caso não haja comprometimento importante de sua validade interna, poderá ser aumentado.[1,2]

EVIDÊNCIAS DO GRADE

Revisões sistemáticas e metanálises devem avaliar as evidências de forma sistemática e reproduzível, reduzindo, assim, o viés de seleção. A metanálise aumenta a precisão dos efeitos estimados, explora as inconsistências entre estudos e, finalmente, fornece inferências aplicáveis a uma gama de pacientes. Metanálises também podem descrever a qualidade das evidências incluindo medidas de precisão. Devemos notar que ensaios clínicos randomizados e estudos observacionais bem conduzidos nem sempre estão disponíveis para orientar todas as questões clínicas. Portanto, uma metanálise comparativa de estudos poderia estar associada a evidências de menor qualidade.[1]

Em particular, o GRADE oferece um sistema para avaliar a força das recomendações. O sistema foi finalmente projetado para revisões e diretrizes que avaliem estratégias ou intervenções de gestão alternativas, o que também pode incluir nenhuma intervenção ou a melhor gestão corrente.[1,2] No entanto, o GRADE é mais do que um sistema de classificação oferecendo um processo transparente e estruturado para a realização das etapas envolvidas no desenvolvimento de recomendações. O GRADE especifica uma abordagem para afiar questões, escolhendo os resultados de

interesse e classificando sua importância, assim como avaliando e incorporando evidências com considerações de valores e preferências dos pacientes para chegar a recomendações específicas.[10] O sistema GRADE (Quadro 5-4) define os quatro graus de evidência – alta, moderada, baixa e muito baixa qualidade – e os níveis de evidência.[1,2]

A qualidade da evidência é um contínuo temporal ou espacial, e qualquer categorização envolve imprevisibilidade. Portanto, a simplificação excessiva a clareza, a transparência e a intuitiva. A compreensão dessas quatro categorias supera essas limitações.[3] Por outro lado, a opinião de especialistas não é um tipo de qualidade de evidência, mas uma interpretação das evidências existentes. Assim sendo, um aconselhamento especializado é quase sempre necessário para integrar e contextualizar evidências de um estudo clínico ou metodológico do ponto de vista. Por exemplo, um estudo randomizado e bem desenhado no seu julgamento poderia fornecer evidências de qualidade contraditória a um ensaio clínico malconduzido. O Quadro 5-5 resume a abordagem GRADE para classificar a qualidade das evidências, que começa com o desenho do estudo (ensaios ou observações de estudos), e, em seguida, aborda cinco razões para possivelmente a qualidade da evidência e três razões para avaliar a qualidade.

Quadro 5-4. Qualidade da Evidência e Evidência do Sistema GRADE

Nível	Definição	Implicações	Fonte de informação
Alto ⊕⊕⊕⊕	Há forte confiança de que o verdadeiro efeito esteja próximo daquele estimado	É improvável que trabalhos adicionais irão modificar a confiança na estimativa de efeito	■ Ensaios clínicos bem delineados, com amostra representativa ■ Em alguns casos, estudos observacionais bem delineados, com achados consistentes
Moderado ⊕⊕⊕○	Há confiança moderada no efeito estimado	Trabalhos futuros poderão modificar a confiança na estimativa de efeito, podendo, inclusive, modificar a estimativa	■ Ensaios clínicos com limitações leves ■ Estudos observacionais bem delineados, com achados consistentes
Baixo ⊕⊕○○	A confiança no efeito é limitada	Trabalhos futuros provavelmente terão um impacto importante	■ Ensaios clínicos com limitações moderadas

		em nossa confiança na estimativa de efeito	■ Estudos observacionais comparativos: coorte e caso-controle
Muito baixo ⊕○○○	A confiança na estimativa de efeito é muito limitada Há importante grau de incerteza nos achados	Qualquer estimativa de efeito é incerta	■ Ensaios clínicos com limitações graves ■ Estudos observacionais comparativos com presença de limitações ■ Estudos observacionais não comparados ■ Opinião de especialistas

Nível	Significado	Fonte dos resultados	Exemplo
A Alta	É muito improvável que trabalhos adicionais irão modificar a confiança na estimativa do efeito	Ensaios clínicos randomizados bem-conduzidos e com achados consistentes. Em algumas situações, estudos observacionais bem-conduzidos, cujos resultados mostram efeitos muito fortes de intervenções terapêuticas que não podem ser explicados por potenciais vieses	Redução de mortalidade com o uso de inibidores da ECA em indivíduos com cardiomiopatia (diversos ensaios clínicos com alto rigor metodológico, em diferentes populações, apresentando consistência nos achados)
B Moderada	Trabalhos futuros poderão modificar nossa confiança na estimativa de efeito, podendo, inclusive, modificar a estimativa	Ensaios clínicos randomizados com importantes problemas na condução, inconsistência nos resultados, achados não dirigidos para o alvo ou desfecho de interesse, imprecisão nas estimativas, vieses de publicação Estudos observacionais, quando relatam benefício forte em delineamento sem viés	Redução de mortalidade cardiovascular com o uso de aspirina em pacientes com insuficiência cardíaca (maioria dos estudos avaliava benefício através de desfechos compostos)

C Baixa	Trabalhos futuros muito provavelmente terão um importante impacto em nossa confiança na estimativa de efeito	Ensaios clínicos randomizados com desfechos substitutos ou outras importantes limitações Estudos observacionais, mais especificamente estudos de coorte e caso-controle	Melhora cognitiva com inibidores da colinesterase em indivíduos com demência (ensaios clínicos com fraco rigor metodológico e inconstância nos resultados)
D Muito baixa	Qualquer estimativa de efeito deve ser vista como incerta	Estudos observacionais não controlados e observações clínicas não sistematizadas (p. ex., relato de casos e serie de casos)	Recuperação de função motora com descompressão cirúrgica do nervo mediano em indivíduos com síndrome do túnel do carpo (evidências escassas na literatura)

Quadro 5-5. GRADE: Avaliação, Desenvolvimento e Avaliação da Classificação de Recomendações. Um Resumo da Abordagem GRADE para Classificar a Qualidade das Evidências

Desenho de estudo	Qualidade inicial	Mais baixo	Superior	Qualidade
Ensaios randomizados	Superior	Limitações no desenho e/ou execução do estudo	Grande magnitude do efeito	Alto
		Inconsistência dos resultados	Gradiente de dose-resposta moderado	Moderado
Estudos observacionais	Mais baixo	Indireção de evidência	Se nenhum efeito observado, todos os confundidores plausíveis Reduzir o efeito demonstrado ou aumentar o efeito	Baixo
		Imprecisão dos resultados Viés de publicação		Muito baixo

O GRADE fornece critérios de qualidade adicionais, da mesma forma que reduz ou aumenta a qualidade das evidências para cada estudo avaliado.[4-12] Limitações metodológicas, como a presença de uma proporção de pacientes perdidos no acompanhamento, falta de cegamento em um estudo randomizado ou vários outros vieses, são responsáveis pela redução da qualidade da evidência.[1,13,14] Em situações em que diferentes critérios levariam a diferentes recomendações, a classificação de evidência clínica deve estar relacionada com o limiar da questão clínica (Quadro 5-6).[13]

Quadro 5-6. GRADE – A Avaliação da Qualidade da Evidência

Qualidade inicial da evidência	***Design* de estudo**	**Abaixar se**	**Maior se**
Alta	Ensaio clínico randomizado, revisão sistemática, metanálise	Limitações do estudo Inconsistência Indireção Impressão Viés de publicação	Magnitude do efeito Gradiente dose-resposta Todos os confundidores plausíveis Ter efeito reduzido
Moderada	Ensaio clínico randomizado, revisão sistemática, metanálise com problemas na condução		
Baixa	Estudo observacional (estudo de coorte, estudo de caso-controle)		
Muito baixa	Qualquer outra evidência (série de casos)		

DA EVIDÊNCIA ÀS RECOMENDAÇÕES

Cada recomendação baseia-se na evidência derivada dos resultados de questões clínicas. Quando a qualidade da evidência difere entre os resultados, o GRADE estabelece que o grau de qualidade das evidências determine a qualidade geral das evidências.[1,2]

Consequentemente, a qualidade geral da evidência seria moderada.[3] Ao usar as recomendações GRADE, os profissionais de saúde têm necessidade de compreender a força da recomendação e da classificação da certeza da evidência. Além disso, informações adicionais cruciais, tais como observações, efeitos completos das alternativas de tratamento, ou considerações sobre valores e preferências e custo, devem estar prontamente acessíveis. As recomendações desenvolvidas com a abordagem GRADE são classificadas como fortes ou fracas. Uma recomendação forte reflete a confiança do painel de especialistas que uma consequência da ação proposta supera ou não os efeitos indesejáveis.[15]

A superação dos efeitos desejáveis de uma intervenção sobre os efeitos indesejáveis reflete a força de uma recomendação. Recomendações fortes significam que a maioria dos pacientes informados escolheria o tratamento recomendado e que os profissionais de saúde podem estruturar suas interações com os

pacientes em tal conformidade.[16] O fraseado usual para recomendações fortes é “recomendamos”. Uma recomendação “fraca” reflete um equilíbrio tênue entre benefícios e desvantagens, incerteza quanto à magnitude das vantagens e desvantagens, mudança ou alta variabilidade nos valores e preferências de pacientes, e ausência de uma justificativa adequada para o custo da intervenção planejada.[15]

Recomendações fracas significam que as escolhas dos pacientes variam de acordo com seus valores e preferências, e os profissionais de saúde devem assegurar que o atendimento ao paciente esteja de acordo com estes valores e preferências.[16] Uma recomendação fraca é geralmente apresentada como “sugerimos” ou “pode ser considerado” (Quadro 5-7). Tais informações podem ser encontradas no resumo das tabelas, que fornecem a taxa para todos os resultados relevantes e os valores numéricos para as estimativas relativas e absolutas do efeito.[15]

Quadro 5-7. GRADE – Classificar os Níveis de Evidência e Força das Recomendações de Acordo com os Parâmetros do GRADE

Grau de recomendação	Recomendação forte	Recomendação moderada	Recomendação fraca	Recomendação negativa
Conclusão de evidência	Benefícios mais fortes que risco	Benefícios maiores que o risco	Benefícios iguais ao risco	Sem benefícios
Alto nível de evidência	Recomendação forte baseada em alto nível de evidência	Recomendação moderada com base no alto nível de evidência	Recomendação fraca com base no alto nível de evidência	Nenhuma recomendação com base no alto nível de evidência
Nível moderado/ baixo de evidência	Recomendação forte baseada em nível moderado/ baixo de evidência	Recomendação moderada baseado em moderado/baixo nível de evidência	Recomendação fraca baseado em moderado/baixo nível de evidência	Nenhuma recomendação baseado em moderado/baixo nível de evidência
Nível muito baixo de evidência	Recomendação forte baseada na opinião de especialistas	Recomendação moderada com base no baixíssimo nível de evidência	Recomendação fraca com base no baixíssimo nível de evidência	Nenhuma recomendação com base no baixíssimo nível de evidência
Relato das recomendações	Nós recomendamos Deveríamos É recomendado/ indicado/útil/ eficaz	Nós sugerimos É provavelmente recomendado Pode ser útil ou eficaz	Nós podemos sugerir Pode ser considerado	Nós não recomendamos Não é útil/eficaz

GRADE é um sistema abrangente, objetivo e robusto que exibe reprodutibilidade suficiente para ser usado todos os dias na prática mesmo por painéis de especialistas com conhecimento relativamente limitado à metodologia de pesquisa.[17] No entanto, o GRADE tem sido desenvolvido para abordar questões sobre gerenciamento alternativo, estratégias, intervenções ou políticas; não foi preparado para perguntas sobre risco ou prognóstico. Preparar uma diretriz implica várias etapas do sistema GRADE. Por esmagador consenso, o GRADE provou seu mérito na avaliação do tratamento e de intervenções além da abordagem clínica, em vez das questões públicas de saúde, mesmo que isso não elimine a necessidade de intervenções; assim como na abordagem clínica, em vez das questões públicas de saúde, mesmo que isso não elimine a necessidade de revisões (Quadro 5-8).

A força de uma recomendação é determinada pelo equilíbrio entre consequências desejáveis e indesejáveis de estratégias alternativas de manejo, qualidade de evidências, variabilidade de valores, e preferências e uso de recursos.[16] Recomenda-se que os especialistas considerem o tipo de evidências que eles estarão avaliando e também a experiência de pesquisa dos avaliadores antes de selecionar uma avaliação crítica e um sistema de classificação. Se a literatura a ser revisada engloba uma mistura mais ampla de metodologias, incluindo análises qualitativas e sociológicas ou evidências baseadas em opinião, bem como pesquisa quantitativa, é recomendado que todos os avaliadores tivessem treinamento em avaliação e classificação de provas usando o sistema a ser empregado para a diretriz (Quadro 5-9).[18]

Quadro 5-8. GRADE – Fatores que Afetam a Força da Recomendação

Fator	**Explicação**
Equilíbrio entre efeitos desejáveis e indesejáveis	Quanto mais significativa a diferença entre os efeitos desejáveis e indesejáveis, maior a garantia de verossimilhança Recomendação firme Quanto mais estreito o gradiente, maior a probabilidade de uma recomendação condicional/fraca
Qualidade de evidência	Quanto maior a qualidade da evidência, maior a probabilidade de uma recomendação substancial
Valores e preferências	Quanto mais valores e preferências variarem, ou quanto maior a incerteza em valores e preferências, maior a verossimilhança garante recomendação condicional/fraca
Custos	Quanto maiores os custos de intervenção

(alocação de recursos)	(maiores os recursos consumidos), menor a probabilidade de uma recomendação substancial

Quadro 5-9. GRADE – Implicações de Recomendações Fortes e Condicionais/Fracas para Diferentes Usuários de Diretrizes

Usuário	**Forte**	**Condicional ou fraca**
Pacientes	A maioria desejaria a ação recomendada. Apenas uma pequena proporção não	A maioria gostaria que a ação fosse sugerida. Muitos não
Profissionais de saúde	A adesão a esta recomendação pode ser usada como critério de qualidade ou indicador de desempenho Ajuda formal à decisão não é necessária para ajudar os indivíduos a tomar decisões consistentes com valores e preferências	As ajudas de decisão facilitam os indivíduos a tomar decisões consistentes com seus valores e preferências
Formuladores de políticas	Recomendação adotada como política na maioria das situações, incluindo para o uso como indicadores de desempenho	Debates substanciais e envolvimento de muitas partes interessadas Os indicadores de desempenho se

		concentram em deliberações adequadas sobre opções de gerenciamento

Pelo contrário, a metodologia GRADE considera aspectos adicionais ao fazer recomendações, mas eles não fazem parte da qualidade do julgamento das evidências. Por outro lado, o GRADE se concentra menos no tipo de estudo e considera múltiplos aspectos dos dados disponíveis ao avaliar a qualidade. No entanto, o sistema GRADE é demorado porque os usuários devem considerar vários aspectos durante a avaliação de qualidade, cada um dos quais aberto a algum grau de subjetividade. Portanto, uma crítica que poderia ser relatada, legitimamente, para o sistema GRADE é que parece que foi desenvolvido por acadêmicos para acadêmicos. Se a maior complexidade do GRADE em comparação com outros métodos resulta no fracasso dos desenvolvedores de diretrizes clínicas em entender o processo, então isso limitará o GRADE de alcançar o objetivo para ser padronizado. Uma solução potencial é produzir métodos alternativos de exibição de dados para diretrizes mais fáceis de usar para os profissionais de saúde, que são os usuários finais. Os profissionais de saúde podem achar as diretrizes menos confusas e mais simples de assimilar se apenas o resultado da avaliação da qualidade foi apresentado, tendo entendido que esta recomendação foi definida após uma avaliação rigorosa usando a abordagem GRADE.[19,20]

CONCLUSÃO

O GRADE é um critério transparente, explícito e abrangente para graduação da qualidade da evidência e separa a classificação da qualidade das evidências das recomendações. No entanto, é um método complexo para não investigadores e profissionais de saúde de forma geral. Além disso, existe um grau de subjetividade que pode interferir nas recomendações e nas diretrizes. Existe um fator não controlado relativo à não publicação de estudos com resultados

negativos ou ausência de resultados que não deve ser negligenciado, muito embora não seja passível de resolução. O sistema GRADE é, apesar das ressalvas, a melhor ferramenta que temos atualmente para avaliar a tomada de condutas baseada em estudos científicos disponíveis na literatura.

REFERÊNCIAS BIBLIOGRÁFICAS

1. Bertolaccini L, Rocco G. The forest of methodology and the writing of evidence-based medicine papers. Eur J Cardiothorac Surg. 2018:54(4):615-621.
2. Diretrizes metodológicas: Sistema GRADE – Manual de graduação da qualidade da evidência e força de recomendação para tomada de decisão em saúde/Ministério da Saúde, Secretaria de Ciência, Tecnologia e Insumos Estratégicos, Departamento de Ciência e Tecnologia. – Brasília: Ministério da Saúde, 2014.
3. Brozek JL, Akl EA, Alonso-Coello P, Lang D, Jaeschke R, Williams JW, et al. Grading quality of evidence and strength of recommendations in clinical practice guidelines. Part 1 of 3. An overview of the GRADE approach and grading quality of evidence about interventions. Allergy. 2009;64(5):669-77.
4. Brozek JL, Akl EA, Compalati E, Kreis J, Terracciano L, Fiocchi A, et al. Grading quality of evidence and strength of recommendations in clinical practice guidelines. Part 3 of 3. The GRADE approach to developing recommendations. Allergy. 2011;66(5):588-95.
5. Andrews J, Guyatt G, Oxman AD, Alderson P, Dahm P, Falck-Ytter Y, et al. GRADE guidelines: 14. Going from evidence to recommendations: the significance and presentation of recommendations. J Clin Epidemiol. 2013;66(7):719-25.
6. Guyatt GH, Oxman AD, Kunz R, Atkins D, Brozek J, Vist G, et al. GRADE guidelines: 2. Framing the question and deciding on important outcomes. J Clin Epidemiol. 2011;64(4):395-400.
7. Guyatt GH, Oxman AD, Vist G, Kunz R, Brozek J, Alonso-Coello P, et al. GRADE guidelines: 4. Rating the quality of evidence–

study limitations (risk of bias). J Clin Epidemiol. 2011;64(4):407-15.
8. Guyatt GH, Oxman AD, Montori V, Vist G, Kunz R, Brozek J, et al. GRADE guidelines: 5. Rating the quality of evidence–publication bias. J Clin Epidemiol. 2011;64(12):1277-82.
9. Balshem H, Helfand M, Schunemann HJ, Oxman AD, Kunz R, Brozek J, et al. GRADE guidelines: 3. Rating the quality of evidence. J Clin Epidemiol. 2011;64(4):401-6.
10. Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, et al. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. J Clin Epidemiol. 2011;64(4):383-94.
11. Puhan MA, Schünemann HJ, Murad MH, Li T, Brignardello-Petersen R, Singh JA, et al. A GRADE Working Group approach for rating the quality of treatment effect estimates from network meta-analysis. BMJ. 2014;349:g5630.
12. Iorio A, Spencer FA, Falavigna M, Alba C, Lang E, Burnand B, et al. Use of GRADE for assessment of evidence about prognosis: rating confidence in estimates of event rates in broad categories of patients. BMJ. 2015;350:h870.
13. Guyatt GH, Oxman AD, Kunz R, Brozek J, Alonso-Coello P, Rind D, et al. GRADE guidelines: 6 Rating the quality of evidence–imprecision. J Clin Epidemiol. 2011;64(12):1283-93.
14. Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, et al. GRADE guidelines: 7. Rating the quality of evidence–inconsistency. J Clin Epidemiol. 2011;64(12):1294-302.
15. Neumann I, Santesso N, Akl EA, Rind DM, Vandvik PO, Alonso-Coello P, et al. A guide for health professionals to interpret and use recommendations in guidelines developed with the GRADE approach. J Clin Epidemiol. 2016;72:45-55.
16. Guyatt GH, Oxman AD, Kunz R, Falck-Ytter Y, Vist GE, Liberati A, et al. Going from evidence to recommendations. BMJ. 2008;336:1049-51.
17. Kumar A, Miladinovic B, Guyatt GH, Schunemann HJ, Djulbegovic B. GRADE guidelines system is reproducible when instructions are clearly operationalized even among the

guidelines panel members with limited experience with GRADE. J Clin Epidemiol. 2016;75:115-18.
18. Baker A, Potter J, Young K, Madan I. The applicability of grading systems for guidelines. J Eval Clin Pract. 2011;17(4):758-62.
19. Cooper NA, Khan KS, Clark TJ. Evidence quality in clinical guidelines: a comparison of two methods. Acta Obstet Gynecol Scand. 2015;94(12):1283-9.
20. Schünemann HJ, Oxman AD, Brozek J, Glasziou P, Jaeschke R, Vist GE, et al. Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. BMJ. 2008;336(7653):1106-10.

REVISÃO GUARDA-CHUVA

CAPÍTULO 6

RESUMO

A revisão guarda-chuva é uma síntese de revisões sistemáticas que permite que os resultados de avaliações pertinentes a uma pergunta de revisão possam ser comparados e contrastados. Na revisão guarda-chuva, este tipo de provas de síntese considera apenas a inclusão do maior nível de evidência, ou seja, revisões sistemáticas e metanálise. O objetivo deste capítulo é descrever a metodologia e os métodos desenvolvidos para a condução de uma revisão guarda-chuva na pesquisa cardiovascular.

Palavras-chave: revisão guarda-chuva; medicina baseada em evidências; metanálise.

INTRODUÇÃO

As revisões sistemáticas (RS) são fundamentais para a prática de cuidados de saúde baseada em evidências, e também fornecem o mais alto nível de evidências para informar a tomada de decisão.[1-4]

As RS oferecem uma informação rigorosa aos tomadores de decisão para utilizarem a melhor evidência disponível para a tomada uma decisão. Devido ao crescente número de RS disponíveis, o próximo passo lógico e adequado tem sido a realização de avaliações das revisões sistemáticas existentes, o que permite que os resultados de análises separadas sejam comparados e

contrastados, proporcionando, assim, uma melhor evidência aos tomadores de decisão.[5,6]

O motivo para a realização de uma revisão guarda-chuva (RGC) é a necessidade de se resumir as provas de múltiplas sínteses de pesquisas e também oferecer uma avaliação rápida dos elementos de prova para tratar uma ampla base de evidências de alta qualidade em um tópico.[7]

Além disso, uma RGC permite uma avaliação sobre tópicos de revisão de perguntas semelhantes, e com isto é possível observar resultados similares de forma independente e chegar a conclusões geralmente semelhantes. O objetivo de uma RGC é proporcionar uma visão global dos resultados para questões particulares ou um fenômeno. Uma RGC pode fornecer um quadro mais amplo de muitos tratamentos e com isto informar de maneira mais útil as diretrizes e a prática clínica quando todas as opções de gerenciamento precisam ser consideradas.[8]

METODOLOGIA DE AVALIAÇÃO DA RGC – RESUMOS DAS EVIDÊNCIAS E SÍNTESE DA PESQUISA

A RGC deve ser precedida *a priori* por um protocolo de revisão por pares, e deve incluir uma pergunta clara, critérios de inclusão em detalhes, um processo de pesquisa estruturado para localizar e selecionar os relevantes estudos existentes, um método para a avaliação crítica dos estudos incluídos, e um processo formal de extração de dados seguido por um meio de resumir e apresentar estes dados. O objetivo principal da uma RGC é fornecer um resumo das existentes sínteses de pesquisas relacionadas a um determinado assunto ou questão, mas não para ressintetizar, como na metanálise ou na metassíntese, os resultados das revisões existentes ou sínteses.[1-3]

OBJETIVOS E CRITÉRIOS DE INCLUSÃO

O objetivo da RGC deve ser claramente indicado. Eles devem ser amplos e devem informar a especificidade da RGC. Os critérios de

inclusão devem indicar a base sobre a qual os estudos disponíveis serão considerados para inclusão, e deverão servir como um guia para o leitor entender claramente o que é proposto pelos autores. Também devem ser a base das decisões em relação aos estudos a serem incluídos durante a fase de seleção dos estudos RGC.[3]

Uma das características únicas de uma RGC é que as unidades de análise, ou os estudos que vão ser incluídos, são avaliados exclusivamente de revisões sistemáticas e metanálises e não de pesquisas primárias ou originais. A apresentação dos critérios de inclusão de uma RGC deve rever a avaliação da eficácia de uma intervenção (população, intervenção, comparador, resultado). Todos os critérios de inclusão apresentados devem refletir com precisão e ser claramente congruentes com o objetivo da revisão e a pergunta.[1-3]

TIPOS DE PARTICIPANTES

As características dos participantes devem ser detalhadas, incluindo idade, sexo, raça, índice de massa corporal e outros critérios de qualificação que devem definir claramente cada grupo.[1-3]

INTERVENÇÕES E FENÔMENOS DE INTERESSE

As intervenções podem ser focadas considerando, por exemplo, apenas o tratamento com drogas, dieta ou exercício, ou a terapia comportamental de um grupo de pacientes. Também podem ser amplas, incluindo tanto a farmacologia como outras intervenções (p. ex., terapia, dispositivos, dieta, exercício, cirurgia). As RGCs que abordam múltiplas intervenções e tratamentos devem definir claramente cada intervenção potencial de interesse.[1-3]

CONTEXTO E CENÁRIO

O contexto deve ser claramente definido e pode incluir, mas não sendo limitado a, uma contrapartida de localização geográfica ou de outros fatores culturais, bem como o sexo e a raça. Em alguns casos, pode também abranger detalhes sobre o cenário específico

de saúde, tais como cuidados intensivos e primários, e a saúde da comunidade.[1-3]

DESFECHOS

Os desfechos de interesse devem ser predefinidos na RGC, os resultados devem ser relevantes para a questão colocada, e as medidas adotadas devem ser incluídas na descrição, como, por exemplo, a utilização de questionários de saúde. Os desfechos substitutos devem ser explicados e apresentados quando existe uma clara associação com paciente e os resultados, bem como se deve relatar os resultados benéficos e adversos.[1-3]

TIPOS DE ESTUDOS

A RGC deverá incluir apenas revisões sistemáticas e metanálises. Nesta seção, a metodologia e a síntese de investigação devem ser consideradas para a inclusão dos estudos na RGC.[1-3]

ESTRATÉGIAS DE BUSCA

A estratégia de busca para uma RGC deve ter como objetivo identificar toda a síntese de pesquisa relevante para a avaliação da questão. Com isto, a busca deve ser abrangente, e os filtros de pesquisa empregados devem ser detalhados e apresentados de forma sequencial em um único apêndice para todos os bancos de dados que foram pesquisados juntamente com as datas de pesquisa e os bancos de dados de citação biomédica.[1-3]

APRESENTAÇÕES DOS RESULTADOS DA RGC

Uma seção introdutória para os resultados da RGC deve começar com uma descrição narrativa do processo de revisão acompanhada de um fluxograma conforme estipulado nos itens de relatório preferido para revisões sistemáticas e metanálises. Esta seção deve fornecer o contexto suficiente para os resultados para que estes apoiem a inclusão da síntese da investigação na RGC. A relevância da síntese da pesquisa deve incluir a questão crítica da RGC.[1-3]

QUALIDADES DA METODOLOGIA

As revisões sistemáticas e metanálises que são elegíveis para inclusão em uma RGC devem ser avaliadas pelas ferramentas disponíveis para avaliar a síntese de pesquisa, o que inclui o AMSTAR e o ROBIS.[9,10]

Recentemente, foi desenvolvida uma lista para a avaliação crítica da RGC, apresentada no Quadro 6-1.[2]

A apresentação dos resultados da avaliação de qualidade deve incluir um resumo descritivo metodológico global da qualidade dos estudos incluídos.

Quadro 6-1. Avaliação Crítica de uma RGC[2]

- É a pergunta de avaliação clara e explícita?
- Foram os critérios de inclusão apropriados para a questão da revisão?
- A estratégia de busca foi apropriada?
- Foram as fontes e os recursos utilizados para procurar os estudos adequados?
- Foram os critérios para avaliar os estudos apropriados?
- Foi a avaliação crítica realizada por dois ou mais revisores independentemente?
- Foram os métodos usados para combinar estudos apropriados?
- Foi a probabilidade de viés de publicação avaliada?
- As recomendações para a política e/ou práticas foram apoiadas por dados reportados?
- Foram as diretrizes específicas e adequadas para nova investigação?

ACHADOS DE UMA RGC

Para minimizar o risco de viés no processo de RGC, deve-se utilizar uma ferramenta de extração de dados padronizados que deve ser empregada por dois revisores independentes para colher os dados de cada estudo. Estas informações devem incluir:[2]

a) Detalhes de citação.
b) Objetivos e critérios de inclusão da revisão.
c) Tipo de avaliação.
d) Detalhes de participantes.
e) Definição e contexto.
f) Número de bancos de dados pelos quais foi feita a busca.
g) Faixa de data de banco de dados em busca.
h) Data da publicação dos estudos incluídos na avaliação, e a informação de cada desfecho de interesse.
i) Vários estudos, tipos de estudos e país origem dos estudos incluídos em cada avaliação.
j) O instrumento utilizado para avaliar os estudos primários e a classificação de sua qualidade.
l) Os resultados relatados que são relevantes para a questão RGC.
m) O método de síntese/análise empregado para sintetizar as provas.
n) Os comentários ou as notas dos autores da RGC que podem informar qualquer inclusão de estudo.

Os detalhes de interesse para a RGC, como a gama de intervenções, fenômenos de interesse, detalhes populacionais ou diferenças de resultados, devem ser extraídos em detalhe. A apresentação dos achados e os resultados devem alinhar-se à questão RGC, e estas devem ser apresentadas de acordo com o Quadro 6-2.[11,12]

Quando as evidências forem qualitativas, os resultados de cada uma delas devem ser apresentados em de acordo com o Quadro 6-3.

RESUMO DA EVIDÊNCIA

O resumo da evidência para os resultados quantitativos e qualitativos são apresentados no Quadro 6-4.

Quadro 6-2. Tabulação dos Resultados Quantitativos para uma RGC[11,12]

Tópico da revisão				
Intervenção	Autor/ano	Número de estudos/ participantes	Achados/resultados	Heterogeneidade

Quadro 6-3. Tabulação dos Resultados Qualitativos para uma RGC[13]

Tópico da revisão		
Fenômeno de interesse/contexto	Achados sintetizados	Detalhes das estratégias

Quadro 6-4. Resumo da Evidência de Síntese para a Pesquisa[13]

Quantitativa		
Intervenção	Autor/ano	Tópico (p. ex., comportamento depressivo)
Quantitativa		
Fenômeno de interesse/contexto	Autor	Achados sintetizados

CONCLUSÃO

Em conclusão, a RGC é um resumo da síntese das revisões sistemáticas sobre um determinado tópico, e deve ser utilizada nos cuidados de saúde baseada em evidências e nas tomadas de decisões.

REFERÊNCIAS BIBLIOGRÁFICAS

1. Biondi-Zoccai G. Umbrella Reviews: Evidence Synthesis with Overviews of Reviews and Meta-Epidemiologic Studies. 1st ed. Switzerland: Springer; 2016;1:1-391.

2. Aromataris E, Fernandez R, Godfrey CM, Holly C, Khalil H, Tungpunkom P. Summarizing systematic reviews: methodological development, conduct and reporting of an umbrella review approach. Int J Evid Based Health. 2015;13:132-140.
3. Aromataris E, Fernandez R, Godfrey C, Holly C, Khalil H, Tungpunkom P. Methodology for JBI umbrella reviews. Joanna Briggs Institute Reviewers' Manual: 2014 edition/Supplement. Australia: The Joanna Briggs Institute; 2014. p. 1-34.
4. Bastian H, Glasziou P, Chalmers I. Seventy-five trials and eleven systematic reviews a day: how will we ever keep up? PLoS Med. 2010;7(9):e1000326.
5. Hartling L, Chisholm A, Thomson D, Dryden DM. A descriptive analysis of overviews of reviews published between 2000 and 2011. PLoS One. 2012;7(11):e49667.
6. Becker LA, Oxman AD. Chapter 22: overviews of reviews. In Higgins JPT, Green S, editors. Cochrane handbook for systematic reviews of interventions, Version 5.1.0 (updated March 2011). The Cochrane Collaboration; 2011.
7. Smith V, Devane D, Begley C, Clarke M. Methodology in conducting a systematic review of systematic reviews of healthcare interventions. BMC Med Res Methodol. 2011;11(1):15.
8. Worswick J, Wayne SC, Bennett R, Fiander M, Mayhew A, Weir MC, et al. Improving quality of care for persons with diabetes: an overview of systematic reviews: what does the evidence tell us? Syst Rev. 2013;2:26-39.
9. Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Med Res Methodol. 2007;7:10.
10. Whiting P, Savović J, Higgins JP, Caldwell DM, Reeves BC, Shea B, et al. ROBIS: A new tool to assess risk of bias in systematic reviews was developed. J Clin Epidemiol. 2016;69:225-34.
11. Kynoch K, Wu CJ, Chang AM. The effectiveness of interventions in the prevention and management of aggressive behaviours in

patients admitted to an acute hospital setting: a systematic review. JBI Library Syst Rev. 2009;7(6):175-233.
12. Forbes D, Forbes S, Morgan DG, Markle-Reid M, Wood J, Culum I. Physical activity programs for people with dementia. Cochrane Database Syst Rev. 2008(3):CD006489.
13. Finfgeld-Connett D. Management of aggression among demented or brain-injured patients. Clin Nurs Res. 2009;18(3):272-87.

MÉTODOS DE ESCORE DE PROPENSÃO

CAPÍTULO 7

RESUMO

Os métodos de escore de propensão (EP) são a probabilidade de um sujeito receber um tratamento condicional em um conjunto de características de base (confundidores), e são usados para comparar pacientes com distribuição similar de fatores de confusão de modo que a diferença nos resultados forneça uma estimativa imparcial do efeito do tratamento. Esta revisão mostra os conceitos básicos dos EP e fornece orientação para a implementação dos métodos de propensão e outros nele baseados, como estratificação, ponderação e ajuste de covariáveis, se tornando um guia prático para o clínico.

Palavras-chave: escore de propensão; ponderação; estratificação.

INTRODUÇÃO

Os ensaios clínicos randomizados são considerados como o estudo mais cientificamente rigoroso para investigar a eficácia de um tratamento, minimizando o viés sistemático. De fato, os sujeitos são aleatoriamente submetidos ao tratamento ou grupo-controle, permitindo, assim, distribuição igualitária entre os dois grupos, balanceando os fatores de confusão medidos e não medidos

(variáveis que influenciam tanto a variável dependente quanto a variável independente causando uma associação espúria, referidas como covariáveis no contexto de regressão).[1]

No entanto, os estudos randomizados podem ser difíceis de serem conduzidos e os estudos observacionais podem fornecer evidências importantes. Em estudos observacionais, os indivíduos nos grupos de tratamento e controle provavelmente diferem entre os fatores de confusão e as diferenças nos resultados podem refletir diferenças nas condições de base, em vez de um efeito real do tratamento.[1]

Combinar cada variável no grupo de tratamento com sujeitos no grupo-controle com confundidores de linha de base comparáveis é uma maneira intuitiva de minimizar confusões em estudos observacionais. No entanto, a correspondência simultânea em poucos confundidores é um processo muito complexo e geralmente resulta em um número muito limitado de correspondências semelhantes. Um método alternativo é a correspondência com base no escore de propensão (EP).[2]

O EP é a probabilidade de um sujeito receber um tratamento T condicional ao conjunto de fatores de confusão (X), e é comumente estimado por meio de regressão logística. O objetivo de estimar o EP é simplificar o processo de correspondência ao reduzir todos os fatores de confusão em um único valor. A correspondência de pacientes com um EP estimado similar cria um equilíbrio aproximado para todos os fatores de confusão, e a diferença nos resultados dentro dos grupos com um PS similar fornece uma estimativa imparcial do efeito do tratamento.[3,4]

A correspondência de EP pode contornar algumas limitações da modelagem de regressão multivariada-padrão[5] (Quadro 7-1), e tem aparecido cada vez mais em pesquisas clínicas.[6] Este capítulo fornece uma orientação para a implementação de métodos baseados em EP para promover transparência e consistência, assim como para facilitar a interpretação dos resultados do estudo.[1]

Quadro 7-1. Vantagens da Correspondência de EP sobre a Regressão MV-padrão

Problema com regressão MV	Comentários	Vantagens da correspondência de EP
Número restrito de fatores de confusão no modelo	No modelo MV, o número de confundidores é limitado pelo número de eventos. Uma regra comum é uma covariável para cada 8-10 eventos. Isso limita a aplicação do MV modelo, particularmente em caso de grande número de confundidores e relativamente baixo número de eventos	Para o cálculo do EP, o número de confundidores usado no modelo PS não é limitado pelo número de eventos resultantes O escore permite que o investigador inclua todos os possíveis fatores de confusão que de outra forma pode não ter sido possível incluir e melhore a eficiência estatística Portanto, o uso do EP pode ser garantido quando o número de confundidores é grande ou o número de resultados é pequeno
Invalidade de o estudo devido a confundir por indicação	Pacientes com contraindicações ao tratamento experimental (ou aqueles com indicações	A correspondência no EP se concentra diretamente nas indicações para o tratamento experimental. Fazer

	absolutas) podem não ter Sujeitos expostos (ou sujeitos não expostos) para estimativa válida de diferenças relativas ou absolutas nos resultados. Esses assuntos geralmente não são reconhecidos com modelagem de resposta convencional e esta pode ser influenciada devido à modificação da medida de efeito ou modelo incorreto	a comparação gráfica dos PSs em sujeitos expostos versus não expostos pode identificar estas áreas de nãosobreposição que são de outro modo difíceis de descrever em uma configuração multivariada com muitos fatores que influenciam as decisões de tratamento
Suposição de modelagem	O modelo de regressão MV baseiase nos pressupostos de modelagem de linearidade entre as covariáveis e o logaritmo natural das chances do resultado	Correspondência pelo EP elimina a hipótese de linearidade entre o EP e os resultados
Design de modelo não separado do resultado da análise	Modelos de regressão MV ajustamse para confundidores modelando relacionamento entre covariáveis e	A correspondência de EP estima o efeito do tratamento por modelagem covariáveis e atribuição de tratamento. Os

	desfecho. Portanto, a especificação do modelo pode ser influenciada pela expectativa do pesquisador para provar a hipótese original	espelhos de correspondência EP são utilizados em um experimento randomizado porque o desenho do estudo (modelo EP e correspondência) é separado do resultado da análise. Isso protege contra suspeitas reais ou viés por parte do pesquisador

MV: multivariável; EP: escore de propensão.

MÉTODOS DE ESCORE DE PROPENSÃO

Existem quatro métodos diferentes baseados em EP:

1. Correspondência: corresponde a um ou mais casos de controle com EP que é (quase) igual ao EP para cada caso de tratamento.
2. Estratificação (subclassificação): divide a amostra em estratos ordenados por classificação. As comparações entre grupos são realizadas dentro de cada estrato.
3. Ponderação: pondera casos pelo inverso do EP. Semelhantemente ao uso de amostragem de pesquisa, pesos são usados para garantir que as amostras são representativas de populações específicas.
4. Ajuste de regressão: isso inclui EPs como uma covariável em um modelo de regressão usado para estimar o efeito do tratamento.

O método EP deve ser escolhido principalmente com base na estimativa de interesse, que depende da questão da pesquisa e da população-alvo.[1]

As estimativas mais comuns são o “efeito médio do tratamento no tratado” (ATT), que é o efeito no grupo de tratamento; e o “efeito de tratamento médio” (ATE), que é o efeito em todos os indivíduos (tratamento e controle). O ATE é mais interessante se todo tratamento potencialmente pode ser oferecido a todos os sujeitos, enquanto o ATT é preferível quando as características do paciente são mais propensas a determinar o tratamento recebido. A correspondência pode estimar apenas o ATT, a ponderação pode estimar os efeitos com base em como os pesos são definidos, a estratificação pode estimar os efeitos com base em como os estratos são ponderados e, por fim, o ajuste de covariáveis pode estimar apenas o efeito marginal, mas não o ATT e nem o ATE.[1]

Ao estimar o efeito do tratamento sobre desfechos binários (*odds ratio*), os resultados de comparação em estimativas têm menos viés do que a estratificação ou o ajuste de covariável. A **probabilidade inversa de ponderação de tratamento (PIPT)** deve ser usada para estimar as diferenças de risco, particularmente quando o interesse é estimar o ATE. Ao estimar o efeito do tratamento nos resultados de tempo até o evento, a correspondência e a PIPT resultam em estimativas menos enviesadas do que a estratificação ou o ajuste de covariável. Recomendamos o uso dos seguintes pacotes R: MatchIt ou não aleatório, não aleatório para estratificação e twang para ponderação.[1]

ETAPAS NA ANÁLISE BASEADA NA PONTUAÇÃO DE PROPENSÃO

A seguir, estão as etapas básicas para remover os efeitos de confusão do efeito do tratamento:

1. Decidir sobre confundidores para que o equilíbrio seja alcançado.
2. Medida (por exemplo, o EP).
3. Condição na medida de distância (usando correspondência, ponderação ou subclassificação).
4. Avaliar o equilíbrio nas covariáveis de interesse, a análise baseada em EP é um processo interativo, e processos e

métodos alternativos baseados em EP devem ser tentados até que uma amostra bem equilibrada seja obtida.
5. Estimar o efeito do tratamento na amostra condicionada.[1]

SELEÇÃO DE CONFUNDIDORES

Os confundidores (X) utilizados para o modelo EP não devem ser influenciados pelo tratamento (T), e eles devem ser medidos (observados antes que T seja dado). A possível explicação para a atribuição do tratamento deve ser fornecida, incluindo a preferência do médico, ou as políticas ou mudanças temporais na prática. Condições preexistentes em unidades de controle para as quais um determinado tratamento não é aplicável são removidas da população do estudo. Uma seleção não parcimoniosa de confundidores é recomendada para reduzir o viés residual. No entanto, a inclusão de muitos fatores de confusão pode reduzir o número de boas correspondências e, portanto, diminuiu a precisão.[1]

Uma abordagem razoável é incluir os fatores de confusão relacionados ao resultado e à atribuição do tratamento, se o tamanho da amostra for grande, e se concentrar em variáveis que se acredita serem fortemente relacionadas ao resultado, se a amostra for pequena.[1]

CÁLCULO DE PONTUAÇÃO DE PROPENSÃO

A grande maioria das aplicações do EP utiliza a regressão logística para estimar a pontuação,[3,4] com a atribuição do tratamento usado como variável dependente e todos os fatores de confusão selecionados forçados como covariáveis EP. A consideração de termos de interação é recomendada.[7] A capacidade preditiva do modelo incluído não deve representar uma limitação na construção do modelo EP. Na verdade, o modelo EP não é empregado para fins inferenciais, mas simplesmente para criar um equilíbrio e, portanto, a prática comum de relatar a estatística-C como uma medida da adequação de uma EP é questionável. Uma estatística-C muito alta pode indicar não sobreposição na distribuição do EP entre indivíduos tratados e não tratados, e sugere uma incapacidade de

fazer comparações entre sujeitos tratados e não tratados. Além disso, uma alta estatística-C não pode ser tomada como prova de que o EP incluía todos os fatores importantes de confusão.[1]

COINCIDINDO

Dois métodos de correspondência comumente selecionados são: a mais próxima correspondência vizinha e a correspondência ideal.[3,4] Vizinho mais próximo depende de um algoritmo guloso que seleciona um participante tratado aleatoriamente, sequencialmente se move através da lista de participantes e coincide com a unidade tratada com a correspondência mais próxima do grupo de comparação. O algoritmo de correspondência ideal minimiza a distância total entre os grupos correspondentes.[1]

Existem várias opções para aumentar a qualidade das correspondências: correspondência com substituição e correspondência com ajuste de pinça. Dentro da correspondência com substituição, um participante de controle pode ser emparelhado várias vezes se o EP dessa pessoa fornecer a correspondência mais próxima à dos participantes de múltiplas intervenções. Combinar com substituição requer que os erros-padrão sejam estimados usando métodos, e estimadores sanduíches como dados não são mais independentes e resulta em perda de precisão.[8,9]

A combinação de pinças usa uma distância pré-especificada dentro da qual as correspondências são consideradas aceitáveis. Se a melhor correspondência estiver fora da distância do cursor, as partidas não serão incluídas no conjunto final. A distância designada é geralmente uma fração de um desvio-padrão do logit do EP (p. ex., 0,20 SD).[7]

As simulações de Monte Carlo mostraram que, quando comparado com outros métodos, o ajuste de calibre resulta em estimativas com menos vieses quando comparado com a correspondência ótima e de vizinhos mais próximos, mostrando melhor desempenho quando avaliado usando-se erro quadrático médio. A correspondência com substituição não tem desempenho superior quando comparada com a comparação de calibradores sem substituição.[10] Comparado com outros métodos, a

correspondência de calibradores resulta em estimativas com menor viés quando comparada com a correspondência ótima e mais próxima e mostra o melhor desempenho quando avaliada usando-se média quadrática erro; a correspondência com a substituição não tem desempenho superior quando comparada com a correspondência de calibradores sem substituição.[11]

Finalmente, o número de unidades de comparação selecionadas para cada unidade tratada deve ser > 1 quando houver poucas boas combinações nos grupos de controle para cada unidade tratada, pois uma maior proporção aumenta a precisão. No entanto, se houver um número limitado de unidades de comparação, uma proporção > 1 poderá ser selecionada como mal correspondida, levando a uma tendência.[11]

ESTRATIFICAÇÃO

A estratificação subclassifica os indivíduos com base nos quantis dos EPs.[12] Os resultados dos indivíduos são então comparados dentro de cada um dos estratos, e um estimador comum do efeito de tratamento é derivado da combinação dos resultados ao longo dos cinco estratos.[1]

Uma prática comum é dividir o EP em cinco estratos. Isto tem mostrado eliminar 90% do viés de confundidores medidos. A estratificação aproxima correspondência sem executar o risco de perder pacientes incomparáveis. Outra vantagem da técnica de estratificação é que permite o cálculo de ambos o ATE e o ATT. As estimativas de efeito específicas do estrato são ponderadas pela proporção de indivíduos que se encontram nesse estrato. Assim, quando a amostra é estratificada em n estratos de tamanho igual, os pesos específicos do estrato de 1/n são comumente usados quando se agrupam os efeitos do tratamento específico do estrato, o que permite estimar o ATE. O uso de pesos específicos do estrato que são iguais à proporção de sujeitos tratados que se encontram dentro de cada estrato permite estimar o ATT.[12]

Uma desvantagem da estratificação é que ela reduz menos os vieses do que os outros métodos, em particular a análise de sobrevivência.[9] Outra desvantagem é a complexidade de agrupar os

efeitos dos estratos (p. ex., o uso do método de Cochran-Mantel-Haenszel).[12]

PIPT

O EP também pode ser usado como pesos inversos nas estimativas do ATE, conhecido como PIPT.[13] O peso de cada participante é calculado usando-se duas variáveis: T (indicador do *status* de tratamento do participante, sendo 0 no braço de controle e 1 no braço de tratamento) e EP de cada participante. O peso (w) do participante [wATE = T/EP + (1 - T)/(1 - EP)] é igual ao inverso da probabilidade de receber o tratamento recebido pelo participante.[1]

Nesta abordagem, as contribuições dos sujeitos do estudo são ponderadas por 1/EP para pacientes experimentais e por 1/(1 - EP) para pacientes de controle. No entanto, um conjunto diferente de pesos permite estimar o ATE no tratado (ATT): wATT: T + EP (1-T)/(1-EP). Os indivíduos tratados recebem um peso de 1. Assim, a amostra tratada está sendo usada como a população de referência para a qual as amostras tratadas e de controle estão sendo padronizadas.[1]

Além disso, para cenários com mais de dois tratamentos, a ponderação de escore de propensão inversa (IEPI) com o EP estimado por meio de modelos generalizados impulsionados pode ser implementada usando-se esses escores para estimar pesos e efeitos causais. As vantagens de usar a IEPI é que ela retém todos os dados do paciente e reduz o viés mais do que a estratificação e o ajuste de covariável.[9]

AJUSTE DE COVARIÁVEL

O EP pode ser usado como uma covariável no ajuste do efeito do tratamento para diferenças de linha de base. Sua vantagem é que o próprio EP pode incluir muitas covariáveis junto com interações; isso permite que o modelo de regressão de covariáveis subsequente seja mais parcimonioso, incluindo apenas as covariáveis relevantes juntamente com a propensão variável de pontuação. No entanto, uma avaliação formal do equilíbrio entre os grupos de tratamento

não é possível. Além disso, produz estimativas mais tendenciosas e suposições errôneas sobre a relação funcional do EP e do resultado (linearidade e risco proporcional), podendo levar diretamente a estimativas enviesadas.[9]

VERIFICANDO O EQUILÍBRIO

A qualidade dos jogos é baseada na comparação da distribuição de fatores de confusão na amostra casada.[14] O uso de testes de hipóteses e valores de P para comparar o equilíbrio não é apropriado porque não há inferências em relação a uma população.

Eles também combinam mudanças no equilíbrio com mudanças no poder estatístico. Os vieses padronizados (também conhecidos como diferença média padronizada) são recomendados para avaliar o equilíbrio das covariáveis entre os dois grupos.[14]

A diferença média padronizada compara a diferença de médias em unidades do desvio-padrão agrupadas.[14] Um valor maior que 0,10 (10%, caso seja reportado como porcentagem) é comumente considerado um índice de desbalanceamento residual. A falta de equilíbrio pode indicar a necessidade de adicionar termos de ordem superior ou não lineares. Os termos de interação também devem ser considerados, em particular entre as covariáveis mais desequilibradas. A análise também pode ser restrita apenas àqueles sujeitos com EP que se sobrepõem aos de outro grupo (suporte comum).[15]

O diagnóstico gráfico pode ser útil para obter uma avaliação rápida do equilíbrio da covariável na presença de muitas covariáveis. O primeiro passo é examinar a distribuição do EP nos grupos original e combinado e a sobreposição do EP usando o histograma espelhado e um gráfico das diferenças padronizadas de médias (nos dá uma visão geral se o equilíbrio é adequado).[1]

ESTIMATIVA DO EFEITO DO TRATAMENTO

Os dados correspondentes devem ser analisados usando-se procedimentos para análises casadas, como testes t pareados para variáveis contínuas, enquanto a regressão logística de teste de

McNemar, logit condicional ou efeito misto (pares combinados como efeito aleatório) pode ser usado para desfechos binários.[16]

Para os resultados de tempo até o evento (sobrevivência), o teste log-rank estratificado, o modelo de Cox estratificado ou o modelo de efeito misto de Cox são necessários.[3] Os dados também podem ser analisados usando-se uma regressão-padrão na amostra casada que inclua um indicador de tratamento e as variáveis usadas no modelo EP (duplo robusto), onde o ajuste de regressão é usado para "limpar" o pequeno desequilíbrio residual da covariância entre os grupos. Os métodos para amostras pareadas fornecem melhores estimativas, e os métodos não pareados podem ser apresentados como análise de sensibilidade.[16]

PONTUAÇÃO DE PROPENSÃO E DADOS AUSENTES

Na presença de dados ausentes, a imputação múltipla pode ser usada para criar conjuntos de dados completos dos quais o EP pode ser estimado.

Existem dois métodos propostos: (1) a média dos EPs após múltipla imputação, seguida por inferência causal; ou (2) inferência causal usando cada conjunto dos EPs das múltiplas imputações seguidas pela média das estimativas causais. É aconselhável incluir o resultado no modelo de imputação.[17]

RELATANDO

Embora o uso dessa abordagem analítica tenha aumentado significativamente na pesquisa clínica, o relato atual é muitas vezes inadequado e ambíguo, e isso resulta em problemas de reprodutibilidade e interpretação do estudo. Para melhorar a consistência e a reprodutibilidade, um conjunto de itens a serem relatados tem sido recomendado.[18] Esses itens devem ser então integrados com as categorias Fortalecimento do Relatório de Estudos Observacionais em Epidemiologia (STROBE) para relatos de estudos observacionais.[19]

EXEMPLO

Um estudo transversal multicêntrico foi realizado em 2015 em 30 províncias do Irã com uma amostra de 4.200 estudantes com idades entre 7-18 anos. Foram realizados exames físicos e testes laboratoriais utilizando protocolos-padrão. A análise foi conduzida com base no EP e a regressão logística condicional foi utilizada para avaliar a associação entre sono curto (menos de 8 h por dia) e o início do sono com a síndrome metabólica (SM) e seus componentes. Os resultados da regressão logística condicional foram relatados como *odds ratios* (OR) e intervalos de confiança de 95% (IC). O EP foi calculado com base no modelo logístico condicional com potenciais variáveis de confusão (idade, sexo, área de vida, comportamentos alimentares saudáveis e não saudáveis, história familiar de doenças crônicas e IMC parental). Dois grupos (com SM e sem SM) foram pareados com base no método de correspondência 1 sem substituição pelo escore. Em seguida, os percentuais de assimetria padronizados (AP) antes e depois da sincronização foram calculados e, em seguida, a média AP foi calculada para todas as variáveis. Devido à correspondência PS, a regressão logística condicional foi utilizada para avaliar a associação entre a duração curta do sono (como variável categórica) e o início do sono (como variável contínua) com SM e seus componentes.[20]

Os resultados da regressão logística condicional foram relatados como *odds ratios* (OR) e intervalos de confiança de 95% (IC). No geral, 3.843 dos participantes completaram a pesquisa. No modelo multivariado, os indivíduos que dormiam menos de 8h por dia tinham uma probabilidade significativamente maior de SM (OR 2,05, IC95% 1,19-3,63) e pressão arterial elevada (PA) (OR 1,46, IC 95% 1,04-2,06). A associação entre a duração curta do sono com outros componentes da SM (incluindo obesidade abdominal, hipertrigliceridemia, hiperglicemia e baixos níveis de lipoproteína de alta densidade) não foi estatisticamente significativa (P > 0,05). Além disso, a associação entre o início do sono com SM e seus componentes também não foi estatisticamente significativa (P >

0,05). A curta duração do sono foi associada ao aumento do risco de SM e PA elevada em crianças e adolescentes.[20]

LIMITAÇÕES

A principal limitação dos métodos EP é sua incapacidade de controlar confundimentos não mensurados. Uma desvantagem da correspondência é um tamanho de amostra muitas vezes substancialmente reduzido porque, para alguns pacientes, as correspondências podem não ser encontradas. Isso pode afetar significativamente as conclusões finais do estudo, que então se aplicam apenas ao subconjunto selecionado de pacientes que poderiam ser correspondidos.[1]

O EP tende a funcionar melhor em amostras maiores e os desequilíbrios significativos de certas covariáveis podem ser inevitáveis, apesar de uma EP ser bem construída e secundária a um pequeno número de observações. Como estudos randomizados, os métodos EP geram um efeito médio e, portanto, não abordam qual tratamento pode ser adequado para um determinado paciente.[1]

CONCLUSÃO

Os métodos EP reduzem um conjunto de fatores de confusão em uma única e intuitiva variável que aperfeiçoa a correspondência e possibilita o ajuste estatístico quando a proporção de eventos para confundidores é baixa. Eles também podem revelar casos em que as populações de pacientes são muito divergentes para fazer comparações significativas.

REFERÊNCIAS BIBLIOGRÁFICAS

1. Benedetto U, Head SJ, Angelini GD, Blackstone EH. Statistical primer: propensity score matching and its alternatives. Eur J Cardiothorac Surg. 2018;53(6):1112-1117.
2. Rubin DB, Thomas N. Matching using estimated propensity scores: relating theory to practice. Biometrics. 1996;52(1):249-64.

3. Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. Multivariate Behav Res. 2011;46(3):399-424.
4. Stuart EA. Matching methods for causal inference: a review and a look forward. Stat Sci. 2010;25(1):1-21.
5. Rubin DB. The design versus the analysis of observational studies for causal effects: parallels with the design of randomized trials. Stat Med. 2007;26(1):20-36.
6. Austin PC. A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003. Stat Med. 2008;27(12):2037-49
7. Brookhart MA, Schneeweiss S, Rothman KJ, Glynn RJ, Avorn J, Stürmer T. Variable selection for propensity score models. Am J Epidemiol. 2006;163(12):1149-56.
8. Austin PC. The performance of different propensity score methods for estimating marginal odds ratios. Stat Med. 2007;26(16):3078-94.
9. Austin PC. The performance of different propensity score methods for estimating marginal hazard ratios. Stat Med. 2013;32(16):2837-49.
10. Westreich D, Cole SR, Funk MJ, Brookhart MA, Sturmer T. The role of the c-statistic in variable selection for propensity score models. Pharmacoepidemiol Drug Saf. 2011;20(3):317-320.
11. Austin PC. A comparison of 12 algorithms for matching on the propensity score. Stat Med. 2014;33(6):1057-69.
12. Cochran WG. The effectiveness of adjustment by subclassification in removing bias in observational studies. Biometrics. 1968;24(2):295-313.
13. McCaffrey DF, Griffin BA, Almirall D, Slaughter ME, Ramchand R, Burgette LF. A tutorial on propensity score estimation for multiple treatments using generalized boosted models. Stat Med. 2013;32(19):3388-414.
14. Austin PC. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity score matched samples. Stat Med. 2009;28(25):3083-107.
15. Blackstone EH. Comparing apples and oranges. J Thorac Cardiovasc Surg. 2002;123(1):8-15.

16. Austin PC. Comparing paired vs non-paired statistical methods of analyses when making inferences about absolute risk reductions in propensity-score matched samples. Stat Med. 2011;30(11):1292-301.
17. Qu Y, Lipkovich I. Propensity score estimation with missing values using a multiple imputation missingness pattern (MIMP) approach. Stat Med. 2009;28(9):1402-14.
18. Yao X, Wang X, Speicher PJ, Hwang ES, Cheng P, Harpole DH, et al. Reporting and guidelines in propensity score analysis: a systematic review of cancer and cancer surgical studies. J Natl Cancer Inst. 2017;109(8).
19. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP; STROBE Initiative. Strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. BMJ. 2007;335(7624):806-8.
20. Hemati Z, Mozafarian N, Heshmat R, Ahadi Z, Motlagh ME, Ziaodini H, et al. Association of sleep duration with metabolic syndrome and its components in children and adolescents; a propensity score-matched analysis: the CASPIAN-V study. Diabetol Metab Syndr. 2018 Nov 3;10:78.

ÍNDICE REMISSIVO

A

B

C

D

E

L

M

R

S

T

V